

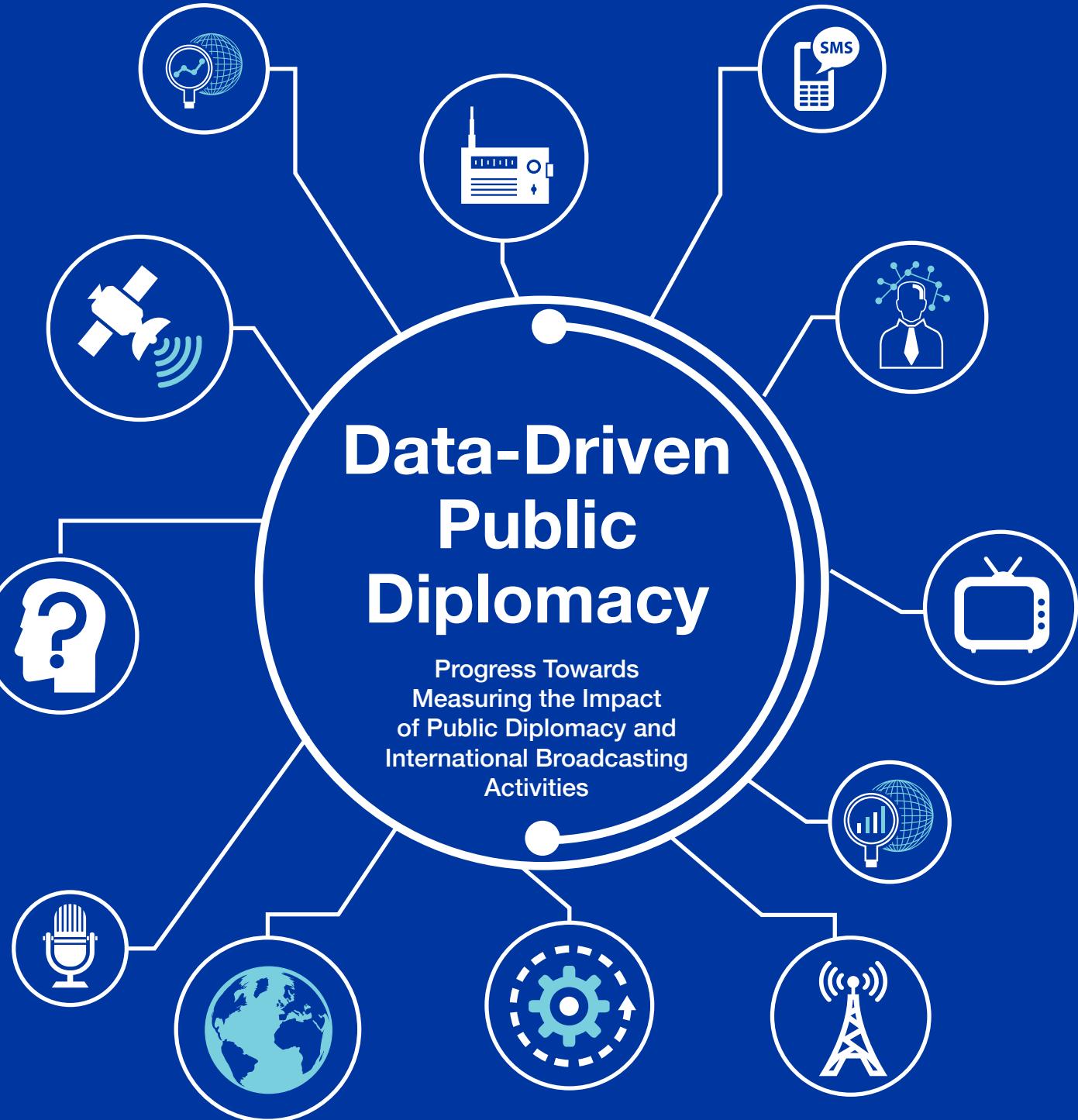


ACPD

United States Advisory Commission
on Public Diplomacy

Data-Driven Public Diplomacy

Progress Towards
Measuring the Impact
of Public Diplomacy and
International Broadcasting
Activities



TRANSMITTAL LETTER

SEPTEMBER 16, 2014

To the President, Congress, Secretary of State and the American People:

The United States Advisory Commission on Public Diplomacy (ACPD), authorized pursuant to Public Law 112-239 [Sec.] 1280(a)-(c), hereby submits a report on U.S. government public diplomacy programs and international broadcasting.

The U.S. Advisory Commission on Public Diplomacy (ACPD) is a bipartisan panel created by Congress in 1948 to formulate and recommend policies and programs to carry out the public diplomacy functions vested in U.S. government entities, and to appraise the effectiveness of those activities. The ACPD was reauthorized in January 2013 to complete a Comprehensive Annual Report on public diplomacy and international broadcasting activities, which will be released in December 2014. It was also mandated to review the impact of public diplomacy and international broadcasting activities.

This report first examines current efforts underway at the State Department and the Broadcasting Board of Governors (BBG) to assess impact through research, analytics and evaluation. It makes suggestions on structures and methodologies needed to make foreign audience research more robust, impact assessment more institutionalized, and feedback loops for strategy and tactics more systematic. The report is based on findings from a six-month study of nearly 100 State Department and BBG research and evaluation documents, in addition to dozens of interviews conducted between February and August 2014 with the staff responsible for them and with users of them. It largely identifies five major areas of change: (1) increased recognition on the part of State Department officials of the importance of research in public diplomacy; (2) movement away from State Department and BBG's risk-averse cultures, which can negatively impact how research data and evaluations are conceived, conducted, reported and used; (3) more consistent strategic approaches in developing and evaluating public diplomacy and international broadcasting activities; (4) increased training in strategic planning, including research and evaluation; and (5) more funding and personnel to conduct more meaningful evaluations at both agencies that can correct the course of programs and activities. Research and evaluation at both agencies is woefully underfunded, falling well below the 5 percent industry standard.

Due to the ACPD's limited resources, we were proud to enlist the support of a distinguished group of scholars who completed the appraisals and helped us make recommendations based on their findings. They include: Dr. Sean Aday, Associate Professor of Media and Public Affairs and International Affairs, The George Washington University; Dr. Amelia Arsenault, Assistant Professor of Communication, Georgia State University; Dr. Matthew Baum, Marvin Kalb Professor of Global Communications, Harvard University; Dr. Kathy Fitzpatrick, Professor of Journalism and Mass Communication, Florida International University; Dr. Craig Hayden, Assistant Professor, American University; Dr. Erik Nisbet, Associate Professor, School of Communication, Department of Political Science, The Ohio State University; Dr. Shawn Powers, Assistant Professor of Communication, Georgia State University; and Dr. Jay Wang, Associate Professor, Annenberg School for Communication and Journalism, University of Southern California. We accept their findings and endorse the recommen-

dations. Dr. Nicholas Cull, Professor of Public Diplomacy, University of Southern California, also provided a historical preface to the report that explains the trajectory of measurement and evaluation at the now-defunct U.S. Information Agency.

Last, ACPD applauds State Department public diplomacy and BBG leadership for their pioneering work in generating audience research, media analytics and program evaluations to produce more meaningful and impactful engagement with foreign audiences for the U.S. government. We admire the tenacity and the talent of the people who conduct this work. We hope this report supports them with specific suggestions for wider systematic change, in addition to methodological fixes that can alter how they operate in their current environments.

Respectfully Submitted,



William J. Hybl, Chairman
(Colorado)



Sim Farar, Vice Chair
(California)



Lyndon L. Olson, Vice Chair
(Texas)



Penne Korth Peacock
(Texas)



Anne Wedner
(Illinois)



Lezlee J. Westine
(Virginia)

CONTRIBUTORS

Sean Aday, Ph.D. Associate Professor of Media and Public Affairs and International Affairs; Director, Institute for Public Diplomacy and Global Communication; Director, M.A. program in Global Communication at The George Washington University

Amelia Arsenault, Ph.D. Assistant Professor of Communication, Georgia State University; Media and Democracy Research Fellow at the Center for Global Communication Studies, University of Pennsylvania Annenberg School for Communication

Matthew Baum, Ph.D. Marvin Kalb Professor of Global Communications at the John F. Kennedy School of Government, Harvard University

Nicholas Cull, Ph.D. Professor of Public Diplomacy and Director of the Master's Program in Public Diplomacy, University of Southern California

Kathy Fitzpatrick, J.D. Professor and Associate Dean of Graduate Programs and Research, School of Journalism and Mass Communication, Florida International University

Craig Hayden, Ph.D. Assistant Professor, School of International Service, American University

Erik Nisbet, Ph.D. Associate Professor, School of Communication, Department of Political Science, and the Mershon Center for International Security Studies, The Ohio State University

Shawn Powers, Ph.D. Assistant Professor of Communication and Associate Director of the Center for International Media Education, Georgia State University

Jay Wang, Ph.D. Associate Professor, Annenberg School for Communication and Journalism, University of Southern California; Director of the USC Center on Public Diplomacy

EDITORS

Katherine Brown, Ph.D. Executive Director, U.S. Advisory Commission on Public Diplomacy

Chris Hensman Senior Advisor, U.S. Advisory Commission on Public Diplomacy

ACKNOWLEDGMENTS

The ACPD would like to profusely thank the nine core contributors to this report: Dr. Amelia Arsenault, Dr. Matthew Baum, Dr. Nicholas Cull, Dr. Kathy Fitzpatrick, Dr. Craig Hayden, Dr. Erik Nisbet, Dr. Shawn Powers and Dr. Jay Wang. We are also grateful to Dr. Emily Metzgar for her contribution to the July 2014 workshop on the subject held in Washington, D.C. The USC Center on Public Diplomacy, under the direction of Dr. Wang, also provided institutional support to the completion of this report, and we thank Dr. Adam Powell for ensuring we had space at the University of California's Washington office for our July 2014 workshop. Our summer analysts, John Pope and Kayli Westling, also made invaluable contributions to helping the research ran smoothly. And we always give thanks to ACPD's Program Support Assistant, Michelle Bowen, who helps make our office run smoothly. Chris Hensman, the ACPD's Senior Advisor, added depth and steadfast pragmatism to the report, while also designing and finalizing the layout for publication.

We are also deeply grateful to the many office directors of the different research and evaluation units and leaders of the public diplomacy bureaus and BBG for being so open to this study, and for sharing with us their reports. This includes, but is not limited to: Bruce Armstrong, William Bell, Roxanne Cabral, Kelu Chao, Vinay Chawla, Tania Chomiak-Salvi, Juliet Dulles, Kim Feinstein, Valerie Fowler, Dr. Jennifer Lambert, Jean Manes, Josh Miller, Dean Olsen, Macon Phillips, Rick Ruth, Bruce Sherman, Dr. Robin Silver, Susan Stevenson, Dina Suggs, Dr. Christopher Toppe, Lynne Weil, and Moira Whelan. They took the time to meet with us repeatedly to discuss the nuance of research designs and the various bureaucratic constraints that they face daily in conducting this work.

TABLE OF CONTENTS

TRANSMITTAL LETTER	2
CONTRIBUTORS	4
PREFACE: EVALUATION AND THE HISTORY OF U.S. PUBLIC DIPLOMACY	7
METHODOLOGY	14
OFFICES REVIEWED	14
INTRODUCTION	16
EXECUTIVE SUMMARY	18
STATE DEPARTMENT: PUBLIC DIPLOMACY POLICY, PLANNING AND RESOURCES OFFICE'S EVALUATION UNIT	23
STATE DEPARTMENT: THE EDUCATIONAL AND CULTURAL AFFAIRS BUREAU'S EVALUATION DIVISION	29
STATE DEPARTMENT: DIGITAL MEDIA EVALUATION IN THE INTERNATIONAL INFORMATION PROGRAMS BUREAU (IIP/ANALYTICS) AND PUBLIC AFFAIRS BUREAU (PA/ODE)	34
BROADCASTING BOARD OF GOVERNORS: RESEARCH AND EVALUATION APPRAISAL	41
PRELIMINARY ASSESSMENT OF THE BBG IMPACT FRAMEWORK	46
CONCLUSIONS	56
SUMMARY OF RECOMMENDATIONS	57

PREFACE: EVALUATION AND THE HISTORY OF U.S. PUBLIC DIPLOMACY

By Nicholas J. Cull¹

By the end of 2003, there was no shortage of issues weighing on the practitioners of America's public diplomacy. The 1999 merger of their former institutional home—the United States Information Agency (USIA)—into the Department of State had been poorly handled. Some questioned the direction of U.S. foreign policy, especially the invasion of Iraq, and many balked at a new emphasis on a particular kind of evaluation of their output summed up in a phrase that rapidly became a favorite of their political masters, "moving the needle." The phrase came from the investigation into U.S. public diplomacy chaired by former Ambassador Edward Djerejian, "Changing Minds, Winning Peace: a new strategic direction for U.S. public diplomacy in the Arab and Muslim world." The older generation of public diplomats resented being asked whether their work "moved the needle" of public opinion towards U.S. foreign policy objectives and saw the approach as reductionist and reflecting a narrow emphasis on the political perspective of the moment. But "moving the needle" was a harbinger of an institutional emphasis on measurement. The discomfort of veteran practitioners in 2003 raises the question of what the previous institutional approach to evaluation had been. Part of the irritation lay in what some perceived as the "reinvention of the wheel." Was there anything in the practices of the past that might illuminate present practice or help understand the culture onto which current evaluation practices are grafted?

Examining the history of evaluation at USIA it is surprising how much was done. The short answer to the "how" question is that U.S. public diplomacy tended to measure output rather than outcome, especially at headquarters. The more complex answer is that the issue of evaluation was just one more aspect of U.S. public diplomacy that depended on the time and place under consideration. Despite the single term public diplomacy the origins of U.S. public diplomacy

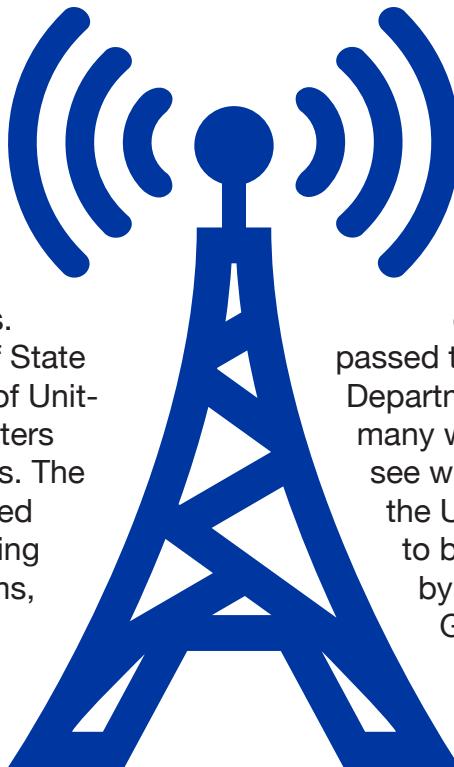
were anything but a uniform experience. U.S. public diplomacy grew from multiple activities housed in multiple agencies. The term public diplomacy arrived only in 1965 at the tail end of the process of institution building and was embraced because USIA itself had a vested interest in establishing a single approach to U.S. public diplomacy and not because one already existed.² U.S. public diplomacy has always included multiple approaches, of which advocacy through press work, cultural and exchange diplomacy, and international broadcasting are the most enduring. Each of these approaches has its own source of credibility, its own professional culture, and its own time-scale. This being so, it is hardly surprising that each had its own approach to evaluation.

The nature of the early history of U.S. public diplomacy weighed against a culture of evaluation taking hold. Evaluation tends to come to the fore when the initiating agency seeks to focus an established range of activities and output. It is seldom on the agenda in a time of crisis when resources are committed hastily in the hope that something will work. Unfortunately, for the first 150 years of its existence U.S. public diplomacy was deployed only as a crisis tool. No one during the American Revolution, Civil War, World War I, or minor crises in between seems to have been overly concerned with exactly which tool of the international communication delivered results. Congress, a major audience for evaluation in modern times, was more concerned to wind down the activity when the crisis was done than to analyze its strengths.

The first steps in building the contemporary structure of U.S. public diplomacy were taken in the early 1930s, when mindful of the rising international challenges the Department of State instituted the so-called Wireless File—a daily compendium of written materials supporting or explaining U.S. foreign policy or political life

assembled or created for journalists overseas. Its success was measured in column inches. The next phase was the creation of the State Department's Department of Cultural Affairs in 1938. This unit focused on mutual educational exchanges between the United States and Latin America. Its founders, such as Ben Cherrington, believed that such work was of value specifically because it was focused on the long-term. The philosophy of Cultural Affairs was always qualitative, though, through the years, the unit developed a wealth of materials to document the value of exchanges.³ Cherrington was much perturbed by pressures from his political masters to evaluate and shape U.S. cultural exchange based on the strategic agenda of America's confrontation with either Nazi Germany or later Soviet Russia.⁴

United States entry into World War II brought a round of new initiatives directed at global publics with a further set of approaches to evaluation. Particularly influential in their impact on the culture of evaluation were the psychological warfare techniques directed against enemy areas by the U.S. Army. These delivered immediate results in the form of individual or even mass surrenders of enemy troops in response to leaflet drops. A generation of senior military figures, most notably Dwight D. Eisenhower, interpreted their own experience as evidence of the over-arching significance of a psychological approach to international relations. The war also saw an expansion of State Department work with a network of United States Information Service centers and cultural attaches at embassies. The Office of War Information developed international programming, including publications and documentary films, but of all the wartime innovations the most significant was Voice of America (VOA). Founded in early 1942, VOA swiftly developed its own evaluation culture.



The shortwave radio station was almost always struggling either for resources or to push back against allegations of program bias and hence maintained a ready supply of positive examples to fire back against critics and help corral resources. There were the expected measurements of output: numbers of broadcast hours, languages put on air, types of programming. In time, VOA developed an evaluation division to monitor the quality of programming going on air. While it was initially hard for VOA to gauge the impact of its broadcasts on target populations, the response of the enemy government was often used as an indicator. From the start of the Cold War, VOA counted Soviet jamming of its signal as a sign of effectiveness. Official anti-VOA street demonstrations in Cold War Eastern Europe and legal crackdowns on radio listening were also seen as badges of effectiveness.

The end of World War II engendered an immense new kind of public diplomacy effort: the re-education of Germany and Japan. This effort included the creation of newspapers, magazines, documentary films, cultural centers and libraries, and a host of other programs. The chief metric of success used to evaluate re-education was the spread of democratic practices. However, the programs were sufficiently stable and sustained and were thus the subject of multiple evaluations for the purpose of fine-tuning output. The effort in West Germany was especially interesting. When the work passed to civilian control under a State Department High Commissioner of Germany work included a survey in 1952 to see which sources of information about the United States that Germans found to be the most credible. The winner by a massive margin was "other Germans who have visited." The High Commissioner responded by investing heavily in exchanges. Educational and leader visits flourished to such an extent

that 10 years later 31 percent of the Bundestag, 53 percent the Bundesrat and 70 percent of the West German cabinet had visited the United States on some sort of exchange program.⁵

The Truman years created a further tier of public diplomacy programs. Each initiative had its own approach and metrics of success including the Fulbright exchanges, expanded press and publicity work at the State Department, and information work associated with the European Recovery Program (Marshall Plan). The Marshall Plan's evaluators tracked public awareness of its work and were delighted by the results. As early as 1947, researchers in Norway reported 94 percent awareness of the basic plan with 74 percent of Norwegians able to answer "technical questions" about the plan's workings.⁶ In 1948, U.S. public diplomacy received its major funding stream through the passage of the Smith-Mundt Act. The act addressed the issue of program evaluation in two ways. The first was the requirement that information materials be available for congressional scrutiny if requested. The second was the creation of the two advisory commissions—the United States Advisory Commission on Information and the United States Advisory Commission on Educational Exchange—that were required to include within their reports to Congress and the Secretary "appraisals... as to the effectiveness of the several programs."⁷ Successive generations of commissioners brought skills relevant for the evaluation of public diplomacy from their backgrounds in the private sector most especially when pioneers of public attitude research, George Gallup and A.C. Nielsen, served on the commission.⁸

As the alphabet soup bubbled away, the need for coordination became apparent. In April 1951, President Truman created an inter-agency Psychological Strategy Board (PSB) whose mission included "evaluation of the national psychological effort."⁹ The system failed to end inter-departmental wrangling. In 1953, President Eisenhower entered office promising to reform U.S. public diplomacy. His response was to create the United States Information Agency (USIA) as

a focal point for the country's overt use of communication in foreign policy. The new agency brought a welcome coherence to the post-war effort. The centerpiece of USIA's practice in the field was an annual planning document known as the "country plan." These detailed documents were developed for each ambassador by the Public Affairs Officer and the associated team at each embassy, with attention to both local and global foreign policy objectives and the dynamics of particular conditions and audiences. The process of writing of the country plan provided an obvious opportunity to evaluate what had gone before and fine-tune the approach for the coming year. Many veterans considered it one of the secrets of USIA's success, though there were some who balked at the scale of the process and the tendency of some of their seniors in post or at headquarters to demand multiple drafts.¹⁰ The performance of individual USIA officers in the field were evaluated both by their area directors in Washington and by the embassy team (project results often figured in those assessments). The double scrutiny struck some as redundant. There was also a team of agency inspectors who would examine procedures, programs, and personnel as required.¹¹

The newly born agency incorporated the network of libraries, embassy press offices, film units, magazines, visitor programs, and Voice of America under its aegis, and its employees were as diverse as their institutional points of origin. USIA also inherited a small group of public opinion analysts from the Department of State's Office of Intelligence Research (OIR), which at USIA became the core of an Office of Research and Intelligence (IRI), founded in 1954. IRI studied both major trends in opinion and evaluated agency tools and programs.¹² Politically important inquiries included an investigation into the VOA Hungarian Service's programming during the Hungarian Rising of 1956, which cleared VOA of inciting rebellion.¹³

In the wider agency, former journalists and publishers worked alongside veterans of public relations and advertising, language teachers,

and humanities academics in an eclectic mix of experience. When called on to demonstrate their effectiveness—as during the compilation of the bi-annual report to Congress—measures cited were diverse, ranging from audience for film or TV programs, the reach of VOA, enrollments in English classes, and the circulation of printed materials. The most consistently scientific approach was that commissioned from outside teams by the administrators of the exchange program (still located at the Department of State). Their multiple studies built a first rate case for the value of exchanges, showing strong correlations between personal experience and positive opinions towards the United States at the end of the period of exchange. Alarm bells in the studies included strong evidence of exchange participants concerns over the issue of race in America.¹⁴

The 1960s were the apogee of USIA's domestic reputation. John F. Kennedy's choice of high profile journalist Edward R. Murrow as agency director leant a season of glamour to the agency. Murrow's greatest response to the evaluation culture at USIA was his memorable observation made in response to pressure from Congress for evidence of effectiveness, "We cannot judge our successes by sales...No cash register rings when a man changes his mind."¹⁵ Even so the agency demanded indications of on-going achievement. The crises of the era—Cuba; Berlin; Birmingham, Ala.—all required attention to international opinion, and the agency could demonstrate a valuable role in each. However, one issue took center stage: public diplomacy in and about the U.S. involvement in Vietnam.¹⁶

The bureaucratic trends associated with the Vietnam War left their mark on USIA. Most especially, the agency inherited the Planning Programming and Budget System (PPBS), which the Secretary of Defense Robert McNamara had brought with him from Ford Motor Company. As agency veteran Barry Fulton recalled:

"USIA selected 'exposures' as the metric for evaluation, and each post was asked

to identify the number of 'exposures' for each product and then produce a summary of cost per exposure. For example, each day a Fulbright grantee spent in the U.S. was 'defined' as one exposure. Each person attending a Dizzy Gillespie concert was defined as one exposure. Each person passing by a window display of a 'Paper Show' [a USIA poster exhibition] was defined as one exposure. So, what headquarters learned from this absurd exercise was that the cost of a Fulbright exposure was many hundreds of dollars, whereas the cost of a paper show exposure was far less than a cent. I hope that decision makers did not make judgments based on the data, but posts were certainly heavily involved in collecting and/or inventing the data for a period, as I recall, of two years."¹⁷

If there were one lesson taught by the 1960s, it was that the best organized, wealthiest, and most ingeniously devised public diplomacy operation in the world couldn't make a bad policy good—as was the case with the intervention in Vietnam—or hide the fact that a society was ripping itself apart. It was equally clear from multiple evaluations that a lot of good work could be done to build goodwill even in the direst political circumstances. An agency evaluation of its documentary film work in 1972 found that posts considered "space," "U.S. life," and "problems and solutions" to be the most valuable themes. U.S. ideology, Vietnam, and non-space science were the least.¹⁸ The findings did not prevent political appointees within the agency applying leverage to try and compel posts to sell the Nixon administration's approach to Vietnam or the rigid anti-communist approach of the era's director, Frank Shakespeare.

By the 1970s, a number of landmark agency programs had been operating long enough to generate some fairly spectacular indicators of long-term success. One favorite of agency boosters was the International Visitor Program. U.S. public diplomats of the era were able to point

to a number of high-profile allies of the country who had been cultivated as rising stars by USIA, most notably Egypt's Anwar Sadat and Britain's Margaret Thatcher. The Carter administration's response to the economic crisis of era included an executive branch-wide evaluation/budgeting procedure borrowed from industry called Zero Based Budgeting (ZBB). At USIA, as Barry Fulton recalled:

“Every program was subject to a ‘what-if’ exercise in which policy makers were asked to describe the consequence of reducing each program by a fixed percentage, 5 percent, 10 percent, 15 percent... This meant ad hoc judgments of ‘effectiveness’ (however defined) by policy makers, and ended in serious cuts, depending in part on the persuasiveness of the ZBB presentations. The *Wall Street Journal* cited USIA as one of the three agencies that had taken the exercise most seriously.”¹⁹

The 1980s saw a revival of the fortunes of USIA. The agency was relevant to the renewed struggle against the Soviet Union and it helped that its director, Charles Z. Wick, was a close friend of President Reagan. Wick had a particular enthusiasm for the communication satellite as a tool of public diplomacy and broke new ground with the launch of WORLDNET a satellite television initiative that included both interactive television broadcasts linking U.S. experts to journalists around the world and a feed of news and feature programming. While the qualitative value of WORLDNET to the posts that hosted interactive events was indisputable, the exact reach of WORLDNET programming was harder to determine. Agency personnel despaired somewhat of Wick's rather incautious claims of the network's reach conflated potential and actual audiences.²⁰ The decade ended in spectacular success. The political changes that gripped Eastern Europe in 1989 delivered multiple examples of the power of public diplomacy. The leaders of revolutions spoke powerfully of the influence of U.S. international broadcasting and analysts noted the

role that exposure to western ideas through exchanges had in laying the foundations for change. Ironically the agency's success in Eastern Europe and the Soviet Union contained the seeds of its decline. USIA had always justified itself as a necessity of the Cold War. Once the Cold War was “won” its political pay-masters saw it as an ideal source of a “peace dividend” budget saving. There were hard times ahead.

The 1990s developed as a time of trial for USIA. In 1992, USIA reorganized to ready itself for the new era of heightened cultural exchange and to preserve resources simultaneously. Innovations included the creation of an Office of Policy and Evaluation within the Educational and Cultural Affairs bureau. The office was small and, as its first director David Wilson later lamented, lacked the funds to commission the kind of in-depth program evaluations from outside academic evaluators that were standard for the U.S. Agency for International Development.²¹ The unit did not have a dynamic reputation. In 1995, the U.S. Advisory Commission on Public Diplomacy raised concerns that it was, in essence, measuring the wrong thing. In 1996, as major budget cuts began to bite, the then-Government Accounting Office (GAO) looked to evaluation as a method to judge which programs to recommend for the axe and ended up criticizing the office for failing to measure the contribution of its exchanges to actual foreign policy goals.²² Around the same time Congress, with the support of Vice President Gore, introduced the Government Performance Results Act (GPRA) of 1993, which created a new system designed to tie expenditure to performance across the executive branch. GPRA imposed a considerable administrative burden on the entire agency by introducing a requirement to generate a hierarchy of goals and deliver annual assessments of the extent to which these had been met. While logical on paper, it proved a poor fit for the agency's work. USIA officers bemoaned the futility of generating both goals and measurements solely to meet the GPRA requirements.

In October 1999, as the bureaucracy struggled

to adapt, the Clinton administration abolished USIA and merged the agency's functions into the Department of State. Suddenly the personal achievements of public diplomacy staff were being evaluated by non-public diplomacy specialists and the entire program was at the mercy of a department with a much more limited exposure to and tradition of program evaluations. America's public diplomats had not found their feet when the United States was faced with the shock of the terrorist attacks of September 11, 2001. Suddenly the U.S. government once again faced the kind of crisis that called for a massive public diplomacy response, at the same time it realized the extent to which Cold War public diplomacy capabilities had been allowed to decline. Attention to evaluation has been part of the road back to effectiveness, beginning with an online data collection system called RESULTS. The need for better evaluation was a theme that surfaced in multiple reports by the GAO, Congressional Research Service, and Defense Science Board and was hammered home in the further applications of GPRA by the Bush and Obama administrations. The Bush administration's Office of Management and Budget launched the "Expectmore.gov" web platform that, in its first year, 2006, rated public diplomacy to be "adequate."²³ The State Department responded in 2007 by developing the Mission Activity Tracker report system (in place of RESULTS) to collect data and a succession of initiatives from the Office of Policy, Planning and Resources (R/PPR) within the Office of the Under Secretary including a 2009 project called Advancing Public Diplomacy's Impact (APDI).²⁴ And so the quest to measure what is at least in part un-measurable continues.

Evaluation has always been part of U.S. public diplomacy. Its most celebrated elements—Fulbright exchanges; VOA's broadcasts; the international visitor program—evolved with close attention to evaluation. However not every program was continuously evaluated and fundamental questions about the relevance of particular programs to shorter-term U.S. foreign policy were sometimes avoided. Some practitioners within

public diplomacy saw their work as above the shifting sands of policy. Evaluation was most often used as a method of fine-tuning or justifying what was already being done and not as a way to make difficult decisions between multiple essentially effective programs. Those in the executive or legislative branches who pressed for more evaluation almost always saw it as a way to cut the budget and never provided extra resources to make good programs even better. The persistence of niche cultures of evaluation mirroring the niche professions within USIA—broadcaster, educator, journalist, diplomat—is also notable. One final challenge to the development of a truly comprehensive and objective culture of evaluation was the persistence of political caprice in U.S. public diplomacy. For every excellent program that flourished because of its demonstrable results there seemed to be a program or approach of questionable value that endured for political reasons. At the heart of U.S. public diplomacy was a glaring fact that belied the lip-service that politicians paid to evaluation: every high-level evaluation of the optimal approach to public diplomacy had stressed the need for public diplomats to be consulted at the formation of foreign policy and not merely called in to spin its execution, yet this has almost never been the case. A certain level of cynicism around the "discovery" of evaluation by successive generations of administrators was only to be expected more especially when those administrators were unaware of the previous work in the field and unprepared to provide the resources necessary to evaluate the work effectively.

In a world of increasing competition for resources, in which the new tools of social media make new kinds of evaluation possible even as the communication environment requires an ever more nuanced approach for an ever more savvy audience, evaluation has an unprecedented significance. It must be part of the DNA of public diplomacy's future. The process of transitioning to the new approach requires dexterity in its application. It is, however, possible to point to traditions of strength in both results and evaluation approaches within U.S. public diplomacy.

America's public diplomats have always understood the need to "move the needle," even as they debated exactly what the needle is measuring, where it is pointing, and the time frame in which its movement should be observed. Properly managed the application of evaluation has the power to ensure that U.S. public diplomacy has a future to match its past.

ENDNOTES

1. This study is based largely on materials that I consulted during the writing of my two volumes on the history of USIA, which in their published forms dealt only in passing with the issue of evaluation. I owe a substantial debt to a number of veterans of U.S. public diplomacy who provided detailed comments on my first draft of this piece, specifically: Len Baldyga, Bob Banks, Don Bishop, Fred Coffey, Barry Fulton, Bob Gosende, Bruce Gregory, Bob Heath, Alan Heil, Frank Hodson, Bud Jacobs, Joe Johnson, Peter J Kovach, Bill Kiehl, Eugene Kopp, Pat Kushlis, Donna Oglesby, Bill Rugh, Mike Schneider, Yale Richmond, Tom Tuch, Dick Virden, and some serving officers who wish to remain anonymous.
2. The author has developed this argument at length in an essay "Roof for a house divided: How U.S. propaganda evolved into public diplomacy" anthologized in Jonathan Auerbach and Russ Castronovo (eds) Oxford Handbook of Propaganda Studies, Oxford University Press: New York, 2014, pp. 131-46.
3. A wide selection of this material survives in the archive of the State Department's Cultural Bureau (CU), which is housed alongside the Fulbright papers at University of Arkansas (hereafter UoA), CU collection, Group XVIII: series 2, Program Evaluation (including Evidence of Effectiveness Reports), boxes 334-336.
4. These pressures are noted by Frank A. Ninkovich, *The Diplomacy of Ideas: U.S. Foreign Policy and Cultural Relations, 1938-1950*. Cambridge University Press: New York, 1981, esp. p. 31.
5. UoA, CU collection, box 332/23 Dept. of State: Mutual Educational and Cultural Exchange activities, background paper, FY 1963, see also 335/17.
6. Cited in Nicholas J. Cull, *The Cold War and the United States Information Agency: American Propaganda and Public Diplomacy, 1945-1989*, Cambridge University Press, New York: 2008, p. 45.
7. For the text of the Smith-Mundt Act see <http://www.state.gov/documents/organization/177574.pdf>.
8. Bruce Gregory, "Public Diplomacy: Sunrise of an Academic Field," *Annals of the American Academy of Political and Social Science*, Vol. 616, March 2008, p. 280. Gregory notes: "For more than half a century no issue has figured more prominently in annual commission reports than recommendations that foreign opinion and media research should inform the formulation and communication of policies."
9. Foreign Relations of the United States 1951, Vol. 1, White House directive, 4 April 1951, p. 58 and note p. 921.
10. I am grateful to USIA veterans Fred Coffey, Hans N. Tuch, Bud Jacobs and Peter Kovach for their recollections of this process.
11. An account of an inspection may be found at Association of Diplomatic Studies and Training (ADST) oral history James P. Thurber (December 1990) online at <http://www.adst.org/OH%20TOCs/Thurber,%20James%20Ptoc.pdf>.
12. For Association of Diplomatic Studies and Training (ADST) oral history including service in IRI see Dorothy Dillon interview (May 1988) on line at <http://www.adst.org/OH%20TOCs/Dillon,%20Dorothy.toc.pdf>.
13. ADST oral history Jock Shirley (November 1989) on line at <http://www.adst.org/OH%20TOCs/Shirley,%20John.toc.pdf> and Henry Loomis (March 1987) http://adst.org/wp-content/uploads/2012/09/Loomis-Henry.VOA_toc_.pdf. A parallel process found some questionable broadcasts had been made by RFE.
14. These studies are conveniently summarized at UoA, CU collection, box 336/15, EDX 2-1, Review of Evaluation Studies, 1960.
15. Alexander Kendrick, *Prime Time: The Life of Edward R. Murrow*, Little Brown and Co., Boston: 1969, p. 490.
16. For a survey of Vietnam work see Nicholas J. Cull, "Selling Vietnam," Jo Fox and David Welch (eds), *Justifying War. Propaganda, Politics and War in the Modern Age*, London, Palgrave, 2012, pp. 288-312.
17. Barry Fulton to author, August 2014.
18. NA RG 306 87.0018, director's subject files, box 34, MTV-films, Janicki (IOR/RM) to Hall (IOR/R), 'Comments on the ISS analysis of worldwide survey of agency films...' 3 May 1972.
19. Barry Fulton to author, August 2014. For the GAO's view of ZBB see the GAO report to Congress, Performance Budgeting: Part initiatives offer insights for GPRA implementation, March 1997, <http://www.gao.gov/archive/1997/ai97046.pdf>.
20. For an example of a tendentious claim see Penny Pagano, "TV goes global on WORLDNET," Los Angeles Times, 2 May 1985, p. C1: "In the last half of 1984, Wick said in an interview, 30 [WORLDNET] programs were seen by 1 billion people."
21. See ADST Oral History: David Wilson (January 2001) <http://www.adst.org/OH%20TOCs/Wilson,%20David%20Michael.toc.pdf> esp. pp. 83-88.
22. General Accounting Office report to Budget Committee, House of Representatives, "US Information Agency: Report for Addressing Possible Budget Reductions." GAO/NSIAD-96-179, September 1996, pp. 37-39.
23. <http://www.whitehouse.gov/sites/default/files/omb/assets/OMB/expectmore/detail/10004622.2006.html>
24. The problems with this system as noted in the US Advisory Commission on Public Diplomacy, *Assessing Public Diplomacy: A Notional Model*, 2010 (a project of the LBJ school, Texas), archived on-line at <http://www.state.gov/documents/organization/149966.pdf>.

METHODOLOGY

This report is based on more than 100 documents provided to us by researchers and evaluators at the Broadcasting Board of Governors (BBG) and in public diplomacy bureaus at the State Department. It aims to dig deep, looking not just at structures and general processes that govern this work, but also at design methodology that could provide richer data to guide future programs and campaigns. We reviewed how these two agencies currently engage in research activities measuring the understanding of key audiences and the impact of public diplomacy and international broadcasting activities. We specifically examined (1) how research informs public diplomacy and international broadcasting activities; (2) how these activities are assessed for impact; (3) the structures and policies in place that allow State Department and BBG officials to collaborate and share available research.

A majority of the documents were from fiscal year 2013, although reports from fiscal year 2014 and some dating back to fiscal year 2007 were also appraised. We also conducted more than a dozen interviews at the State Department and at the BBG, meeting with senior managers, research directors, and relevant program staff to understand how research is used to develop information and public diplomacy activities and to evaluate their impact. This work was done from February 2014 through August 2014.

OFFICES REVIEWED

Department of State: The Public Diplomacy and Public Affairs Cone was created in 1999 after the merger between the U.S. Information Agency and the State Department. Its mission is to “support the achievement of U.S. foreign policy goals and objectives, advance national interests, and enhance national security by informing and influencing foreign publics and by expanding and strengthening the relationship between the people and Government of the United States and citizens of the rest of the world.”¹

In July 2014, ACPD convened in Washington the project’s core contributors to discuss the trends they identified in research design across these offices and possibilities for structural change, especially at the State Department. The executive summary and key recommendations reflect the findings from that workshop.

At the State Department especially, the work appraised was based mainly on Washington-directed activity. This report does not look at various field measurement and evaluation activities carried out at embassies. We reviewed documentation pertaining to the scope and quality of conducted research in four offices in the Public Diplomacy and Public Affairs Cone (R): the Educational and Cultural Affairs (ECA) Evaluation Division, the International Information Programs (IIP) Audience Research and Measurement Office, the Public Affairs (PA) Office of Digital Engagement, and the Policy Planning and Resources Office’s (R/PPR) Evaluation Unit. [Note: While R/PPR also manages measurement and strategic planning tools like the Mission Activity Tracker (MAT), this report did not look at them in-depth.] At the Broadcasting Board of Governors, it focused on the International Broadcasting Bureau’s (IBB) Office of Performance Review, and the work of its contractor, Gallup, in addition to the BBG’s Office of Strategy and Development.

The Evaluation Unit within the Policy Planning and Resources Office (R/PPR) works to assess the impact of U.S. public diplomacy and to evaluate select programs. In addition to the Advancing Public Diplomacy’s Impact (APDI) report, the team conducts evaluations that gauge the long-term impact of certain programs that inform, engage and influence foreign citizens.

The Evaluation Division within the Educational and Cultural Affairs (ECA) Bureau examines a variety of cultural and educational programs through short “Performance Management Results” reports and longer field evaluations. The core goal of ECA is to enhance mutual understanding, further programmatic objectives (e.g., language teaching), and improve participants’ professional development (e.g., new skills, career development). Evaluations are conducted through assessments, including online surveys, field work (e.g., interviews, focus groups), and document analysis.

The Analytics Office within the International Information Programs (IIP) Bureau is intended to develop audience research and optimize content for IIP programs and campaigns through mostly online tools (e.g., technographics, audience segmentation, message testing). In order to create a feedback loop between program and campaign managers, it measures the content resonance within audiences, and works to identify online conversations and behaviors.

The Office of Digital Engagement (ODE) in the Public Affairs (PA) Bureau maintains the State Department’s core social media content. As part of this practice, they use a variety of commercial tools to assess whether messaging is receiving the expected level of online attention and to evaluate different tactics for spreading information through social media. This data is fed into weekly and monthly reports to PA leadership and reported to State Department leadership. It is done to assess the efficiency of messaging and

improve future work. Because there is no central monitoring and evaluation team in the PA bureau, this office was reviewed along with IIP/ARM due to its analogous digital media outreach activities.

Broadcasting Board of Governors (BBG): The mission of the BBG is distinct from other parts of the State Department’s public diplomacy activities in that its primary objective is not to persuade attitudes regarding the United States or its policies, but to “inform, engage and connect people around the world in support of freedom and democracy.”² It is, however, strategically aligned with broader U.S. foreign policy goals.

The Office of Performance Review and Office of Strategy and Development at the Broadcasting Board of Governors (BBG) were created in 2012. The Office of Performance Review oversees surveys on audience size and media usage, focus groups with topics that are selected by a given language service, and evaluations of individual broadcast programs. The data from this research are used to design programs, analyze BBG’s competition, provide the basis for performance reporting, target specific audiences, and determine if the news on BBG programs is considered credible. The Office of Strategy and Development uses this research to help BBG staff and management evaluate their programs and make research-based decisions about changes. While the BBG conducts a variety of research products, the focus of this appraisal was on reports completed by its contractor Gallup and on the new BBG Impact Framework.

ENDNOTES

1. Under Secretary for Public Diplomacy and Public Affairs. U.S. Department of State. <http://www.state.gov/r/>

INTRODUCTION

In one of his first communications with the public diplomacy community, Under Secretary of State for Public Diplomacy and Public Affairs Richard Stengel, cited the need for “a new 21st century tool kit for public diplomacy.” An important part of that new tool kit will be valid and reliable measurement tools that help to evaluate the efficiency and effectiveness of public diplomacy efforts, their impact on foreign publics’ views of and attitudes toward the United States, and their influence in advancing U.S. foreign policy goals.

Public diplomacy is a vital dimension of American diplomacy. Given the fast proliferation of non-state actors who are shaping the international system this century, it has never been more pertinent to our national security strategy. Often, public diplomacy officers are under pressure not just to produce immediate outputs, but also immediately to demonstrate their results. Yet public diplomacy, like traditional diplomacy, is a long game. Impact measurement takes rigorous and consistent data collection, pre-, mid-, and post-activity, for extended analysis. It also requires funding, experienced specialists, discipline, and a commitment by leadership.

Over the last two decades, governments and multilateral organizations around the world have expressed a greater interest in program and performance monitoring and evaluation.¹ Calls for these efforts toward U.S. public diplomacy efforts expanded in earnest in the mid-2000s after a series of high-profile reports commissioned in the wake of September 11 highlighted the need to “move the needle forward” in terms of enhancing America’s image abroad.²

Public diplomacy leadership began to feel greater pressure to demonstrate the efficacy of short-term initiatives and programs. In 2006, the Office of Management and Budget (OMB) had requested an increase in performance measurement efforts government-wide. The Government Accountability Office’s (GAO) 2007 report,

“Strategic Planning Efforts Have Improved, but Agencies Face Significant Implementation Challenges,” also cited a deficit in State Department research in informing strategic communications campaigns compared to the Department of Defense and the U.S. Agency for International Development’s efforts.³ With the exception of the State Department’s Intelligence and Research Bureau (INR) audience polling capacity, a carry-over function from the U.S. Information Agency, the GAO cited an overall “lack of actionable, in-depth research available” to public diplomacy officers, which discouraged them from identifying objectives based on research. The lack of complex audience research, GAO reported, drives officers to make educated guesses or “gut decisions” on how to communicate, engage and develop relationships with those audiences.⁴

Since then, the State Department’s Public Diplomacy and Public Affairs cone has made considerable progress in institutionalizing both audience research and evaluation in its bureaus. There is no deficit of daily media monitoring, and there has been a significant shift toward creating better tools and recruiting specialists who can produce actionable data. Under the direction of Under Secretary of State Karen Hughes, more mechanisms to track public diplomacy spending and measure aggregated impact were developed with the Mission Activity Tracker (MAT) and the Advancing Public Diplomacy’s Impact (APDI) report, respectively. Digital media outreach also became institutionalized within the Public Affairs Bureau, eventually leading to the use of commercial analytics tools to track message pick-up. In 2010, under the direction of Under Secretary Judith McHale, the International Information Programs Bureau’s (IIP) Office of Audience Research and Measurement (ARM), which was recently renamed the Analytics Office (IIP/Analytics), was created to produce knowledge on audiences and measure the impact of various information campaigns via online tools.

The BBG has been an interagency leader in measurement and evaluation. While the agency's mission to "inform, engage and connect people around the world in support of freedom and democracy" is distinct from State Department public diplomacy activities in the sense that it does not aim to persuade attitudes regarding the U.S. or its policies, it is strategically aligned with foreign policy goals such as serving as a model of a free press and supporting its role in free, democratic societies, and fostering greater understanding and engaging audiences. The Office of Strategy and Development and the Office of Performance Review in the International Broadcasting Bureau (IBB) were established in 2012 to understand audiences and impact by measuring the efficacy of its programming in achieving some or all of their objectives in the target countries within which it operates. Their deep dives on audience research in countries vital for U.S. foreign policy have been particularly valuable to the inter-agency community, and to outside stakeholders.

Research and evaluation of programs is a fairly young concept throughout the State Department; the Public Diplomacy cone has been a pioneer in trying to measure activities' outcomes. Created in 2006, the Budget and Planning Office at the State Department in the Management cone, was mandated to look at the budget process and has tried to build the capacity to integrate strategy, plans, and programs across the department. Until then, the Educational and

Cultural Affairs Bureau's (ECA) Evaluation Division, formed in 1999 after the merger with the U.S. Information Agency, was the only dedicated office to review short-term performance results and long-term evaluations. It has provided guidance and template models for the Budget and Planning Office in trying to understand the impact of hundreds of State Department activities.

As the de facto leader for research and evaluation at the State Department, the Public Diplomacy and Public Affairs cone has the opportunity to continue to lead by example. Based on interviews and discussions with a dozen State Department officials responsible for public diplomacy measurement, and a review of available public diplomacy and public affairs evaluation reports, it appears that a commitment has been made to making impact measurement a central component of public diplomacy efforts. The recent movement toward more data-driven planning and measurement is a positive shift in public diplomacy operations, and the officials involved demonstrated significant expertise.

Those involved in international broadcasting and public diplomacy research and evaluation are impressive both for their deep loyalty to U.S. diplomacy generally and the need for measurement specifically. Given this commitment, we feel it is an opportune time to rethink how actionable research is created and impact evaluation is conducted and to strengthen the public diplomacy tool kit in order to ensure its maximum utility for public and traditional diplomacy leadership.

ENDNOTES

1. In the US, a series of presidential directives—e.g., the 1993 Government Performance and Results Act (GPAR), the 2002 President's Management Agenda and Program Assessment Rating Tool (PART), the 2009 OMB Program Evaluation Initiative, and the 2011 High Priority Performance Goals (HPPG) Initiative—catalyzed offices and agencies throughout government to adopt evaluation techniques designed to illustrate efficacy.

2. See for example: Djerejian, Edward. Changing Minds Winning Peace: A New Strategic Direction for Public Diplomacy in the Arab &

Muslim World. Report Submitted to the Committee on Appropriations, House of Representatives. Vol. Chaired by Edward Djerejian. Washington D.C.: Advisory Group on Public Diplomacy for the Arab and Muslim World, 2003.

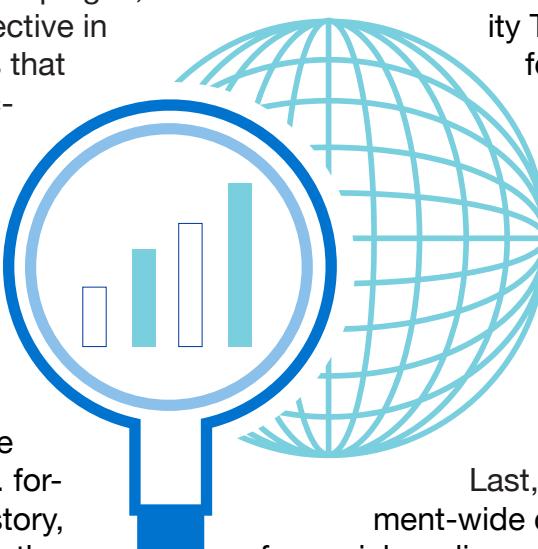
3. [Strategic Planning Efforts Have Improved, but Agencies Face Significant Implementation Challenges](#); GAO-07-795T: Published: Apr 26, 2007. Publicly Released: Apr 26, 2007.

4. Ibid.

EXECUTIVE SUMMARY

State Department public diplomacy leaders and officials at the BBG face several structural constraints in producing consistent, sound research to inform and evaluate campaigns and programs. This includes a lack of time, qualified staff and funds to conduct thoughtful, long-term evaluations; in addition to risk-averse cultures, which can often lead officials to misinterpret setbacks as failures and therefore fail to report them. At the State Department, laws such as the Privacy Act of 1974 and the Paperwork Reduction Act of 1980 are major impediments to conducting timely and useful evaluation and analytical information on- and offline.¹ As a result, the current research and evaluation systems in place often seem to justify programs, campaigns, and budgets. They could be more effective in providing internal feedback loops that could contribute to course corrections and improvements in public diplomacy strategies and tactics.

How does one determine impact? Too often the difference between outputs and outcomes can be conflated. Given the multitude of factors at play that cause foreign audiences to react to U.S. foreign policy—e.g., psychology, history, politics, culture, religion, and the others that shape a worldview—it is likely that a public diplomacy or broadcasting activity or campaign contributes to an outcome rather than directly causing it. But public diplomacy's and international broadcasting's impact can be measured through knowledge, understanding, and trust of the United States, discovered via the compilation of best practices, longitudinal panel studies, comparative studies, and historical case studies. In-depth research also helps to identify and develop culturally appropriate programs and messages, and the proper way to employ them. In addition, it can provide necessary information to eliminate programs and campaigns that are not efficient or impactful.



Improved research and evaluation is dependent largely on structural and organizational change, some of which is already underway at the State Department and at the BBG. In the ECA Bureau, for instance, more robust alumni programs will allow for increased longitudinal panel studies with the same people over time to gauge how perceptions of America, and their relationship with America, change with time. Encouragingly, the recent move by IIP leadership to ensure the re-branded Analytics Office is part of campaign design will help set appropriate goals and metrics at the beginning, therefore creating more informed campaigns at the outset, as well as better post-hoc analysis. The improvements R/PPR is making to the Mission Activity Tracker (MAT) will also allow for better baseline data in the sense that it will be less burdensome for Public Affairs Officers to self-report information, and its creation of a Public Diplomacy Implementation Plan will hopefully support public diplomacy officers with strategic planning.

Last, its deployment of a department-wide content management system for social media accounts at U.S. embassies worldwide will enable better coordination of efforts in digital engagement, and, potentially make for sustainable procedures for pre-and post-communication analytical efforts.

The BBG has focused much time in research and evaluation from its inception with the United States International Broadcasting Act of 1994, which established the Broadcasting Board of Governors. The law mandated a “reliable research capacity” to build programming “based on reliable information about its potential audience.”² ACPD fully endorses the reorganization of existing research capabilities to create a new IBB Office of Research and Assessment, which

will substantially improve the strategic focus and efficiency of research efforts by overseeing the entire research and digital infrastructure throughout BBG, including the introduction of new information intelligence products and tools. Its movement toward an “Impact Framework” to measure the outcomes of its media activities by creating informed publics, establishing sustainable local media, and improving government accountability in the countries it reaches is also an encouraging move.

Below, we present our key recommendations based on the recurring findings in the four chapters that examine the State Department public diplomacy and BBG offices. First, we review suggested modifications to current structural and organizational impediments, especially at the State Department. Second, we review opportunities to modify research designs so that the reports provide a more actionable data and a more effective feedback loop for policy makers. We recognize that some of these design fixes depend on increased specialists, time and funding, therefore necessitating the structural and organizational change.

STRUCTURAL AND ORGANIZATIONAL CHANGE

Most of the recommendations below, apply to both agencies except where noted.

Create a Director of Research and Evaluation Position and Expand the Evaluation Unit in Policy, Planning and Resources (R/PPR): Research and evaluation for public diplomacy at the State Department needs more strategic leadership. This position, established within the Policy Planning and Resources Office, should regularly design and advise on standardized research questions, methodologies, and procedures that directly link practice to strategy and foreign policy objectives. This office would give more organizational legitimacy and authority to research, advocate for researchers’ needs, and prioritize research activities in ways that reflect strategic short-, middle-, and long-term objec-

tives. We recommend that the Director report directly to the Under Secretary and be able to inform s/he, and other State Department leadership, of knowledge in a tangible and interpretable form. The Director of Research would also support research coordination and consistency within the “R” bureau [IIP, ECA, PA and the Center for Strategic Counterterrorism Communications (CSCC)]³ in developing systematic and longitudinal methods. S/he would also be able to conceptualize the questions based on the needs of leadership, match the right mix of methodology to the program, determine the level intra- (e.g., Intelligence and Research Bureau (INR); Democracy, Human Rights and Labor Bureau (DRL)) and inter-agency (e.g., BBG, Department of Defense) research coordination, and determine when the work should be outsourced or completed in-house.

Support Evaluation Staff with More Expertise: The staffs tasked with conducting research and evaluations at both agencies are by and large well trained, dedicated to their tasks, and eager to implement reforms. Yet they are too few of them to expand the range and scope of evaluation techniques. While some contracting is necessary to maintain impartiality, not all of this work should be outsourced. More in-house specialists who can proactively conceptualize research and employ the right methodologies are essential for improving the work.

Increase Funding for Research and Evaluation: While evaluation and measurement officials recognize the need to employ a broad range of methods and participants, and to expand research products, they do not have adequate funding to pursue it. According to the Evaluation Roundtable, an association of evaluation professionals, the standard for research and evaluation in philanthropies and foundations should be 5 percent.⁴ In fiscal year 2014, \$8.8 million of the BBG’s \$726.5 million budget, or 1.2 percent, went toward research and evaluation. This percentage is decreasing in fiscal year 2015 to 0.7 percent (\$4.7 million of \$721.26 million). The State Department’s Educational and Cultural

Affairs Bureau, had less than .25 percent of the \$568.5 million fiscal year 2013 budget allocated toward evaluation. Increasing funding at both agencies so that these units get closer to at least the 4 percent industry norm would allow for more robust and constructive research and evaluation processes to drive decisions on what activity would best support U.S. foreign policy goals.

Review Further The Privacy Act of 1974: The ACPD will conduct a follow-up review of the Privacy Act of 1974 with legal experts and provide separate recommendations based on their analysis. The law currently roadblocks audience research and analytics in the sense that, according to State Department lawyers' interpretation of it, influential figures cannot be identified with online analytics tools, despite the fact that those figures opt to broadcast information publicly using commercial platforms like Twitter. Anonymized information can still contribute to the creation and appraisal of programs and campaigns, yet its value is limited.

Provide State Department Research Waiver for the Paperwork Reduction Act of 1980: Public diplomacy is a vital dimension of national security and research is critical to its success. This law limits the State Department's ability to conduct measurement research in a timely fashion, as research officials must receive a waiver for each study that seeks to enlist more than 10 people. This bottleneck is particularly relevant with respect to the need for benchmark and time-series studies that would reveal public diplomacy's impact over time. The INR Bureau, a member of the U.S. intelligence community, has a blanket waiver.⁵ While not a part of the intelligence community, we feel that public diplomacy research offices should also receive one.

Improve Inter- and Intra-Agency Cooperation and Data Sharing: Evaluation units within the Department of State share data and collaborate on an ad hoc or informal basis. There appears to be no systematic way that evaluations are distributed, stored, or solicited. The need for

timely sharing of information was recognized in the June 2013 State Department Office of Inspector General Report on the IIP Bureau (see Informal Recommendation #2, page 15). Department-wide, the 2007 GAO report also recommended a more holistic look at audiences based on polls, studies, focus groups and in-depth media analysis.⁶ Clear lines of authority to access data that is being gathered by other government units (e.g. Open Source Center), as well audience data collected by third parties, should be established.

Support a Risk-Taking Culture That Allows for Public Diplomacy and International Broadcasting Setbacks: As is the case with almost all bureaucracies, suggestions of limited or negative outcomes may inhibit future funding and administrative support. This creates a climate that inhibits realistic evaluations, and evaluations in general. In the current environment, it is hard to imagine how critical, forward looking research designs could be implemented given existing cultures of fear and risk-aversion. State Department and BBG leadership should reward and encourage honest and balanced evaluations and encourage the admission of setbacks for stronger programming.

Establish Guidance and Training on Research and Evaluation: BBG and State Department managers need written guidance and training on how to read and interpret evaluation findings and should be encouraged to seek out evaluations of previous or complementary research to use actionable information to change programs, and not just validate their work. A program at the Foreign Service Institute on how to identify and integrate basic research and evaluation into A-100 classes for new foreign services officers, not just public diplomacy officers, would help significantly. These skills should also be reinforced through advanced training.

Create ACPD Subcommittee on Research and Evaluation: Dependent on a Memorandum of Understanding with State Department public diplomacy and BBG leadership, the U.S. Advi-

sory Commission on Public Diplomacy should create a sub-committee on research and evaluation to review State Department and BBG research agendas, methodologies and interpretations once a quarter. It would be comprised of selected academics, market researchers, and research professionals from private organizations. The sub-committee's objective would be to provide objective feedback early into the research to ensure the methodology is rigorous and the findings are rich. They would also help establish a set of achievable goals for this research, emphasize the employment of research in strategic planning, and serve as an additional validity check on the suggestions that emerge in the reports.

METHODS AND RESEARCH DESIGN

The six key recommendations below are based on the recurring research design issues in the State Department and BBG offices examined and are not wholly dependent on increased funding and personnel. We suggest that the research and evaluation directors aim to include these suggestions in their work whenever possible.

Acknowledge Research Limitations in Products:

The inherent challenges associated with measuring the long-term impact of public diplomacy are well-known. For example, how do you measure something that didn't happen—e.g., someone doesn't become a terrorist or a reporter doesn't promote anti-American sentiments through national media—even if such non-actions are partially the result of public diplomacy initiatives? The inability to show that public diplomacy efforts are the direct and proximate causes of such effects hinders the ability of public diplomacy officials to provide convincing data to policymakers and others on the value of public diplomacy to U.S. national interests. Thus, this reality should be acknowledged where appropriate.

Increase Integration of Data into Strategy and Program Development:

Evaluations are

only useful if the goals and objectives are clearly defined; clearly defined goals and outcomes are critical at every level of authority. The International Information Programs (IIP) leadership is already doing this: the Office of Analytics is now located with the campaigns and foreign policy offices of the bureau, a change that promises to better link research and evaluation to campaign design and implementation. We applaud this, and encourage further movement toward establishing data-driven goals as the first and most critical phase of public diplomacy planning rather than the final phase. This would help efforts be more information-driven and strategic, and help to ensure that defined objectives are measurable.

Create More Disaggregated Data for Better Feedback Loops:

Aggregated data without context or nuance has very limited value for State Department leadership and may be misleading. Disaggregated data by demographics (e.g., gender, age, economic status) or other targeted criteria, in addition to participants/respondents' specific experience with U.S. public diplomacy, will provide better feedback for the types of programs and experiences that resonate in different countries and regions. This is especially pertinent to the State Department's Advancing Public Diplomacy Impact Report (APDI), which is an aggregated benchmarking study and is used on a limited basis by stakeholders in the State Department and externally.

Use More Comparative Data and/or Analysis to Determine Impact:

At the State Department and BBG, comparative data is either not used enough to show impact (or lack thereof) or it can be used incorrectly. In some cases, comparisons made to illustrate certain trends were based on incongruous variables and could be used in a more effective and precise manner. More comparative data at all of the State Department's public diplomacy evaluation offices and at the BBG would be clarifying for where U.S. programs rank, helping to benchmark outputs and outcomes vis-a-vis the United States' closest allies and competitors.

Provide More Contextual Data to Determine Impact: Rarely in the State Department's or the BBG's reports is contextual data provided to understand why changes in foreign public sentiment or behavior do or do not occur. Public diplomacy activities are often but one factor of many that may determine policy success. It is more likely that a public diplomacy activity or campaign contributed to an outcome in a secondary manner rather than as a primary cause and thus a lack of the "big picture" may provide

misleading conclusions about public diplomacy's impact.

Highlight Negative Findings for Couse Correction: Evaluations should be written in a balanced manner that highlights the successes and *failures* of particular campaigns and activities. Research units need the authority to make such guidance, and leadership must encourage analytical products to be seen as constructive rather than punitive.



ENDNOTES

1. The Privacy Act of 1974 establishes a code of information practices to govern the collection, maintenance, use, and dissemination of information about individuals that is maintained in systems of records by federal agencies: <http://www.gpo.gov/fdsys/pkg/USCODE-2012-title5/pdf/USCODE-2012-title5-partI-chap5-subchapII-sec552a.pdf>; The Paperwork Reduction Act of 1980 was created to limit reduce the amount of paperwork burden the federal government imposes on private businesses and citizens: <http://www.gpo.gov/fdsys/pkg/PLAW-104publ13/html/PLAW-104publ13.htm>.

2. U.S. International Broadcasting Act of 1994.

3. At the time of this study, the Center for Strategic Counterterrorism Communications was in the early stages of creating an analytics and/or

evaluation capacity and was therefore not appraised.

4. Evaluation Round Table.

5. Intelligence and Research Bureau, U.S. Department of State. <http://www.state.gov/s/inr/>

6. Specifically, it stated that "interagency protocols regarding the sharing of audience research information, including establishing a forum that would bring audience research staff together on a regular basis to discuss plans and concerns across all topics of interest" while also developing "an electronic clearinghouse of U.S. government audience research that could be accessed by staff throughout State, USAID, DOD, and BBG, including BBG grantees."

STATE DEPARTMENT: PUBLIC DIPLOMACY POLICY, PLANNING AND RESOURCES OFFICE'S THE EVALUATION UNIT

By Sean Aday Ph.D., Kathy Fitzpatrick J.D., and Jay Wang Ph.D.

Understanding the long-term impact of public diplomacy activities worldwide is essential for strategy and program development. Yet compiling a comprehensive report on it is an enormous task. First published in 2009, the Advancing Public Diplomacy's Impact (APDI) report is the largest evaluation document that the U.S. Department of State produces. It was originally designed in 2007 to be a "comprehensive performance assessment of U.S. Department of State global public diplomacy (PD) efforts" and provide a global yardstick for U.S. performance on engaging foreign publics.

The study employs a quasi-experimental design of focus groups and surveys to attempt to compare responses from participants and alumni of U.S. public diplomacy programs—and/or people on the Public Affairs Section's contact list—to the responses of those who have no contact with the local U.S. Embassy in seven countries. It is intended to be a benchmarking study to assess if public diplomacy activities, writ large, impact target audiences. The Evaluation Unit within the Policy, Planning and Resources Office in the Public Diplomacy and Public Affairs cone (R/PPR),¹ leads this study. The unit also conducts program evaluations (See: Sidebar on The Evaluation of American Centers). The fieldwork for its third iteration is currently being completed by an independent contractor; this appraisal therefore applies to its 2011 report.

In response to the Office of Management and Budget's (OMB) government-wide push in 2006 to assess and improve program performance through research and evaluations, and its concerns about public diplomacy performance, APDI was originally created to collect benchmarking data in aggregated form. Its findings have since been used to respond to data requests from the

OMB, in addition to the Government Accountability Office (GAO), the State Department's Budget and Planning Office, the Congressional Budget Justification for public diplomacy funding, and other oversight agencies. In 2011, the report included 2,800 participants in seven countries and cost \$3 million. (The 2013 report has reportedly included 1,800 interviews in six countries and cost \$1.2 million.)

Such a large-scale study could provide rich insight for policy makers. However, its current design to present aggregated data and its overwhelmingly positive findings do not illuminate how public diplomacy programs are, or are not, impactful over time. There are some questions about whether this type of assessment provides the most valuable data for strategic planning, program development, and foreign policy advancement.

After reviewing the 2009 and 2011 reports in-depth and speaking with the Evaluation Unit team in R/PPR, we recommend that public diplomacy leadership establish a new research design that moves away from collecting primarily aggregated data; increases sub-level analysis for richer insight; connects the research design to U.S. foreign policy goals; focuses on countries and participants critical to U.S. foreign policy goals; supplies greater context of country, regional and global trends; lessens its reliance on self-reported data; and provides constructive criticism. Completing a report that provides actionable data for government officials and Congress will also require fundamental structural and organizational change, which we address more in the opening of this report.

BACKGROUND

This appraisal focuses mainly on the 2011 APDI report, which focuses on seven countries—Bangladesh, Brazil, Cambodia, Indonesia, Kenya, Morocco, and Turkey—to measure the aggregate impact of public diplomacy efforts abroad. It claims to find evidence that “conclusively show(s) that the global portfolio of public diplomacy programs is having its desired effect.” Specifically:

- All 10 performance measures were positive (in expected direction and levels).
- Compared to non-participants, participants in past public diplomacy programs in all seven countries showed higher favorability toward the United States, greater understanding of and favorability to its policies, and more positive opinions about U.S. actions in the world.
- Past programs were more knowledgeable of U.S. foreign policy.
- Past participants also shared their knowledge and opinions about the United States in small group, interpersonal communications and more broadly in speeches and traditional and social media outlets.

The report has six key strengths. First, non-participants are included as a comparison group for data analysis and the study design goes to great lengths to find an appropriate comparison group for its former public diplomacy program participants. Second, comparative studies could reveal specific areas needing attention. Third, the multiple-method approach (e.g., survey, focus groups) provides richer data than a single study. Fourth, face-to-face interviews in local languages can provide more accurate findings since misunderstandings related to study design could be addressed on the spot. Fifth, the training of fieldwork specialists appears to be com-

prehensive. Sixth, the research is conducted by an outside firm, which is generally viewed to be more credible.

From our conversations with key stakeholders and the three-person team within the Policy, Planning and Resources Office (R/PPR), we understand that the purpose of the APDI study appears to be limited to an aggregate analysis of measurement data. Indeed, it seems that a lot of effort and resources go into this report but it's not clear that much happens after it is submitted within the State Department, and to Congress and oversight agencies. There appears to be little discussion between Congress, State and OMB about it after its submission, and the report doesn't appear to be utilized much internally within the State Department. With this much time, effort, and resources devoted to a report with so much over time potential, it should be better utilized.

APPRAISAL OF CURRENT METHODOLOGY AND INTERPRETATION

The overall design of the study seems well thought out if one is interested in making aggregate claims. However, we provide the following recommendations to change the current design to create more actionable data for the decision-makers in Washington, and for the various U.S. embassy officials who help coordinate the study worldwide.

RE-DESIGN THE REPORT TO FOCUS ON DISAGGREGATED DATA

The data provides an aggregated view of public diplomacy influences rather than program-specific results. At a fundamental level, this mandate to make an aggregate level analysis seems problematic. Gleaning information in terms of best practices from such a broad level of analysis is difficult. That said, in our conversation with key figures in R/PPR involved in the APDI, it seems that this is more an issue of carrying out instructions. They have been tasked with doing

an aggregate analysis, so that's what they did. On a limited basis, the evaluation team provides some disaggregated data at the country-level to supplement the report.

Still, for this report, and future ones like it, to assess the impact of America's public diplomacy programs and to find ways to improve and/or replicate them, a more disaggregated approach that allows us to know which programs were more effective and why, in addition to an overall assessment, would be necessary. It would also be helpful to know more details about the projects' past participants (e.g., rural/urban, number of participants, etc.) than the generalized categories offered (e.g., cultural), and which programs work better in which parts of the world, if possible.

Recommendation: Re-design the study to focus less on aggregate findings, so that it can provide the foreign policy community with more detailed and analytical information in linking public diplomacy programs with policy objectives. Field cases studies would support efforts to better understand why and how public diplomacy works—or doesn't work—in achieving desired outcomes with various populations in various parts of the world.

CONNECT U.S. FOREIGN POLICY GOALS WITH RESEARCH DESIGN

The study revealed some keen insights that could be very useful in developing future programs, as well as informing foreign policy. For example, the differences in gender are striking. It appears, however, that this data is not shared for such purposes. The remaining information is "nice to know" but not actionable in terms of informing and guiding public diplomacy strategies. The measurement is also not linked to specific objectives of public diplomacy programs. As a result, it is difficult to determine the impact of specific public diplomacy activities on changes in knowledge, attitudes and behaviors. Changing the study's design to understand why

and how attitudes are formed would help in the development and advancement of ongoing and future programs. Of course, there could be other influences as well on knowledge, attitudes and behaviors such as global and national media and education.

Recommendation: Re-design the future study so that it allows for more "actionable data" and a better feedback loop for policy. This includes describing the insights gained from this study with other public diplomacy practitioners and policy makers so that they might be used in the development and advancement of ongoing and new public diplomacy initiatives and in informing U.S. foreign policy. An expansion of the Evaluation Unit staff would also support the report's redesign and, ultimately, its wider dissemination within the Department and to Congress, oversight agencies, and to stakeholders.

FOCUS ON KEY COUNTRIES AND PARTICIPANTS FOR U.S. FOREIGN POLICY GOALS

When describing the demographics of survey participants, it would also be illuminating to know if those were the target audiences of public diplomacy programs. It is not clear that the countries and participants were key publics important to the State Department's ability to accomplish its mission. Rather, the survey participants seemingly were chosen for reasons unrelated to the strategic goals of public diplomacy. The study focuses on elites and acknowledges that it therefore precludes generalization to the larger population of "ordinary citizens," yet it should strive to include non-elites to understand the broader effect of public diplomacy. It should also say outright if elites were selected because they are the target audience for foreign policy objectives. There should also be a guiding principle for why some countries will be included (e.g., importance to U.S. foreign policy goals). In the upcoming 2014 report, for instance, Jamaica is included in the study, the reason for it would need to be made clear.

Recommendation: Focus on key publics rather than convenient populations and include non-elite publics in future evaluations in an effort to gauge the broader impact of public diplomacy on foreign populations. Currently the non-participant group is primarily determined by demographics, which has quite significant limitations.

INCREASE SUB-LEVEL ANALYSIS FOR RICHER INSIGHT

The report presents mostly data without context or explanatory analysis. Even with the current aggregate design, there seems to be a missed opportunity in not conducting more sub-analysis on the countries repeated in the 2009 and 2011 study. The third iteration of the study currently underway misses the opportunity to return to some of these countries to provide even richer data over time, although we understand concerns about exhausting a sample too soon. One of the key challenges for any information campaign, including those related to public diplomacy, is the endurance of effects. Variations in timing of respondents' participation in various public diplomacy programs could skew results and impact the value of data gathered. For example, the report notes that the majority of participants have had "prolonged exposure and multiple experiences with American PD." Yet, there appears to be no evaluation of how both timing and types of programs, as well as number of programs in which respondents participated, might influence the results. In some countries, participants were much higher on education and other key demographic variables than non-participants, calling into question the value of the comparison group. To that end, the analysis would do well to dig deeper and compare not just those with "multiple" versus single exposure to public diplomacy programs, but how many exposures and how long ago they were, as well as types of exposure. With a redesign, the report could be made stronger with an analysis of why America's public diplomacy programs are

allegedly working in the ways claimed by the study, in addition to why we see differences on some measures between countries, or between subgroups.

Recommendation: Provide more consistency, or clarification and justification of the findings. Suggest including "recency" (based on participant segments of program experiences) and "intensity" (in terms of the nature and amount of exposure to U.S. public diplomacy) in the analysis. Of course, one realizes it is not easy to recruit valid comparison groups in field settings, but this should be aimed for whenever possible and should be made more explicit in the report's analysis.

PROVIDE STRONGER CONCEPTUAL GROUNDING FOR "KNOWLEDGE, INTEREST, AND UNDERSTANDING"

The central constructs of "knowledge, interest, and understanding" of the United States in the research design needs to have stronger conceptual grounding and operationalization. The report seems to conflate the concepts of participants' "interest, knowledge, and understanding" of the United States. The question simply asks respondents for their personal sense of how interested or knowledgeable they are, or how well they think they understand concepts, about the United States. This is problematic because it doesn't adequately measure their underlying concepts. If you cannot establish a baseline understanding of each participant's experience and impressions of the United States, then the interpretation of the data needs to be better contextualized and qualified.

Recommendation: Contextualize and qualify the interpretation of the data that discusses participants' "knowledge, interest, and understanding" of the United States. More specific questions included in the interview schedule could measure subjects' actual knowledge and understanding of the United States. If possi-

ble, establish a basic set of questions of “knowledge, interest, and understanding” for every program participant before they start a public diplomacy program with the United States. As a pilot, this could potentially be incorporated into the performance management results the Educational and Cultural Affairs Bureau (ECA) completes and be disaggregated by country, and then further disaggregated by demographics.

SUPPLY GREATER CONTEXT OF COUNTRY, REGIONAL AND GLOBAL TRENDS TO THE ACTIVITY BEING MEASURED

The presentation of the APDI data needs to be more contextual and thorough and capture the differences in context—geographic, political, etc.—that could influence participants’ attitudes and behaviors toward the United States. The report should include information about the local situations specific to public diplomacy challenges, including an understanding of the local media and governance systems (whether they were facing similar or different challenges at the broad level). In addition, it would help to know the total number of participants in public diplomacy programs in each location, and the nature and scope of the programs in respective countries. Time differences in terms of when participants participated, as described above, also could provide a richer understanding of program impact.

Recommendation: Incorporate greater context of the political, geopolitical, economic, and sociological dynamics affecting the country. The analysis of the data should also consider both the timing and types of programs, as well as the number of programs in which respondents participated. The impact of other influences on participants’ views, such as media consumption, and educational backgrounds, could also be incorporated.

ENCOURAGE CONSTRUCTIVE CRITICISM THROUGH EVALUATION

The fact that the outcome on virtually every measure was strongly positive runs the risk of creating an impression among skeptics that the goal of the report isn’t to better understand U.S. public diplomacy efforts in order to improve and replicate them, but rather to simply placate legislators and justify budgets. Specifically, when the report said “The outcomes for all 10 of the public diplomacy performance measures were positive, that is, in the expected direction and at the expected level,” it begs the question as to what these “expected levels” were. Were they set before the public diplomacy programs? There were some variances among the countries, for instance, in favorability ratings. Were they expected? More clarity about what these expected levels were, and how they were set, would blunt some of these potential criticisms and provide more confidence that the overwhelmingly favorable outcomes are meaningful.

Recommendation: Provide results that are supported by the research design and findings. In the 2011 analysis, for example, there is a need to qualify such statements as “The higher favorability of participants toward U.S. global actions and policies in particular demonstrates without doubt that PD is making strides in achieving its mission of undergirding U.S. foreign policy initiatives and furthering U.S. national interests.” The research design and findings didn’t support such a strong claim.

CONCLUSION

We are encouraged to see an emphasis on measuring the long-term impact of public diplomacy on foreign publics. However, this effort is limited by the reporting requirements placed on researchers. Rather than attempt to “show that public diplomacy works,” the focus should be on better understanding whether and how public diplomacy works—or doesn’t work—over time.

If the Advancing Public Diplomacy Impact study is to become a true benchmarking exercise that will contribute to the development of more effective public diplomacy programs, then both a new study design, increased funding and more personnel support for the Evaluation Unit will be required. Providing wider access of the rich

dataset to academics for further analysis may also generate both strategic and tactical insight for practitioners. Such changes would contribute to the development of a “culture of research” in public diplomacy that would help to advance a more strategic and more effective U.S. public diplomacy operation.

BEST PRACTICE: THE EVALUATION OF AMERICAN CENTERS

Completed in 2014, the Evaluation of American Centers was designed specifically to review the performance of these U.S. government platforms overseas. Since, the study has provided valuable insights for future planning as the Department of State places increasing emphasis on American Centers and their role in achieving public diplomacy objectives. While the report primarily measures output and processes rather than outcomes and impact, as per its instructions, it is exemplary in establishing a feedback loop for public diplomacy strategy.

Knowing more about “how American Centers’ current programming expands engagement with foreign audiences” provides a good foundation both for identifying gaps in current activities and developing future activities. The measurement of “what visitors gain from the experience,” however, appears to be limited. We know little about how foreign publics’ experiences with American Centers changed their knowledge, attitudes and/or behavior on matters important to the achievement of public diplomacy objectives. One exception was survey questions related to what visitors learned about American culture, society, and values as a result of their visit.

The Office of American Spaces within the International Information Programs (IIP) Bureau has used the report to change how they support these platforms abroad. They have prioritized a list of the 10 most strategically important spaces in each of the department’s six regions, as determined by foreign policy goals. A significant amount of IIP’s support funding for upgrades in technology, design, and content will go to these 60 spaces to ensure that they exemplify the best of American modernity and attract new audiences and act as a hub for alumni activity. IIP is also working with the Smithsonian Institution on the Model American Spaces Program to create these model designs.

Moving forward, appraisals of these new American Centers activities should continue to link its efforts to overall public diplomacy goals and develop a research-based strategic plan for each center. The plan should identify key publics (including inactive, or non-users of American Centers) and the public diplomacy impact objectives for each key public.

ENDNOTES

1. The R/PPR office also supports evaluation and strategy development with tools like, but not limited to, the Mission Activity Tracker (MAT) and the Public Diplomacy Implementation Plan (PDIP).

STATE DEPARTMENT: THE EDUCATIONAL AND CULTURAL AFFAIRS BUREAU'S EVALUATION DIVISION

By Kathy Fitzpatrick, J.D., Jay Wang, Ph.D.

The activities of the Educational and Cultural Affairs Bureau (ECA) are widely recognized as the most important and most successful public diplomacy efforts in the State Department. They represent the foundation of the Fulbright Hayes Act of 1961, the purpose of which is to

“Enable the government of the United States to increase mutual understanding between the people of the United States and the people of other countries by means of educational and cultural exchange; to strengthen the ties which unite us with other nations by demonstrating the educational and cultural interests, developments, and achievements of the people of the United States and other nations, and the contributions being made toward a peaceful and more fruitful life for people throughout the world; to promote international cooperation for educational and cultural advancement; and thus to assist in the development of friendly, sympathetic, and peaceful relations between the United States and the other countries of the world.”¹¹

The effectiveness of exchanges conducted by ECA in reaching long-term goals is the most difficult and most expensive to measure. Yet its \$1.3 million expenditures in fiscal year 2013 made up less than .25 percent of the Educational and Cultural Exchange Programs budget. As the oldest evaluation unit within the State Department, ECA has provided much of the direction for research and impact assessment for diplomatic activities writ large through logic models and step-by-step guides. This work is laudable given the financial and bureaucratic constraints.

We reviewed more than a dozen of short- and long-term evaluation documents from 2002 to 2013. We make seven recommendations for

how to adjust the research design: (1) connect program objectives with research design; (2) separate short-term from long-term goals; (3) avoid reports that rely on self-evaluation data; (4) supply greater context of country, regional and global trends; (5) encourage constructive criticism through evaluations; (6) clarify descriptions of research processes; and (7) distinguish between what's inferred versus what is directly assessed or observed.

BACKGROUND

The Educational and Cultural Affairs Evaluation Division has existed since the merger between the U.S. Information Agency and the Department of State in 1999 (the U.S. Information Agency had an audit unit, which later became the evaluation unit). The purpose of the unit is twofold: to understand performance management of various ECA programs and to assess the long-term impact of select programs through evaluations. This is distinct from program management, which seeks to ensure the efficient use of federal funds.

ECA officials appear committed to using available resources to supplement anecdotal information with more quantifiable data. The types of engagement measured include the short-term International Visitor Leadership Program and longer exchanges, such as various programs under Fulbright and the Youth Exchange and Study Programs. They do so mainly through short-term studies via the ECA Performance Measurement Initiative, which has included, over the last 15 years, 60,000 to 70,000 respondents who were surveyed before and after their U.S.-sponsored programs. The ECA evaluation unit also commissions roughly three long-term evaluations per year on select programs, which the ECA leadership requests. These evaluations look at pro-

grams that are relatively new, those that relate to foreign policy, and those that are priorities for the Under Secretary. The ECA evaluation unit designs the methodology and the contractors implement it. The evaluations rely on surveys, interviews, focus groups, and document analysis.

One State Department official interviewed for this report said that ECA programs “are not strategic communication campaigns,” suggesting a lack of strategic planning in ECA programming (or perhaps a limited interpretation of strategic communication campaigns as efforts to manipulate). This thinking appears in contrast to the fact that officials in this area clearly recognize the value of benchmark data in program development.

APPRAISAL OF CURRENT METHODOLOGY AND INTERPRETATION

Reports reviewed for this paper are publicly available on Evaluation Division’s website: <http://eca.state.gov/impact/evaluation-eca>. They include the long-term evaluations of the English Language Specialist Program (2014), the Fulbright English Teaching Assistant Program (2014), SportsUnited Programs (2013), the Edward R. Murrow Program for Journalists Citizen Exchange Program (2012), the Youth Exchange and Study Program (2009), the English Access Microscholarship Program (2007), the School Connectivity Program (2007), the Building Respect through Internet Dialogue and Global Education Program (2007), and the Fulbright Visiting Scholar Program (2002). The assessment also included reviews of the short-term Performance Measurement Reports between 2011 and 2013 for the TechWoman Program, the Global Undergraduate Exchange Program in Eurasia and Central Asia, the TechGirls Program, the Professional Fellows Program, the Youth Ambassador Program, the Study of the United States Institutes Women’s Leadership Program, and the International Visitor Leadership Program.

The evaluations did a very good job capturing the elements of the program and the activities of the beneficiaries. At the same time, anal-

yses were often conducted at the aggregate level, creating a disconnect between program objectives and measurement. Additionally, we identified problems associated with a reliance on self-reported data, a lack of context, and a focus on positive results. Our key findings are discussed below, along with recommendations for improvement.

CONNECT PROGRAM AND FOREIGN POLICY OBJECTIVES WITH RESEARCH DESIGN

The stated objectives of the programs reviewed were not specifically defined or linked to measurement, which can result in a partial assessment. There was also some confusion regarding the aims of the various programs and how they link to overall ECA goals. For example, “mutual understanding” is cited in the programs reviewed both as a program goal and an area for evaluation. One of them, the Youth Exchange and Study Program, “aims to promote mutual understanding and respect between the people of the United States and the participating partner countries.” However, there is no real measure of mutual understanding and no indication of how “mutuality” is defined or achieved in any of the evaluations. If we are to use the term mutuality, it is necessary to clearly define the term and its operationalization, and connect the specialists’ cultural learning with how they helped to increase foreign participants’ understanding of America. One exception was the evaluation for the Fulbright English Teaching Assistant Program, which attempted to measure the impact of American teaching assistants’ experiences on their own cross-cultural knowledge and global perspective.² The measures about “knowledge” and “understanding” used in these various studies seem to have been developed ad hoc. They need to be verified for validity and reliability to enhance analytical rigor. The focus clearly is on “their” understanding of the United States and not “our” understanding of their nations in ECA programming.³

Recommendation: Develop specific research-based goals and objectives that

link programs to foreign policy aims. A set of standard questions/measurements about “knowledge” and “understanding” of the U.S. would also help ensure validity and cross-study comparison.

SEPARATE SHORT-TERM FROM LONG-TERM GOALS

Many of ECA programs are conducted on a regular basis, which provides great opportunity for research/evaluation to help set goals. For instance, if in Year 1 we accomplished this much “mutual understanding,” can we do better in Year 2 if we set goals based on program assessments? It is important to make evaluation more strategic. It is not simply about assessing what we did; rather it is about finding ways to maximize impact. One way to better determine impact is to more effectively differentiate between long-term objectives (e.g., relationship building, contributing to policy change) and short-term objectives (e.g., enhanced awareness, attitude change, image improvement, increased mutual understanding, behavior change).

Recommendation: Distinguish short-term and long-term objectives in program plans. Short-term objectives, as they are in Performance Management Result surveys, are more easily measured and could contribute to program planning and development. Yet they are based on performance results rather than outcomes or impact, therefore limiting insights for program development. However, there is an opportunity to tie them into long-term evaluations. Also, more specificity would provide better data to show actual impacts versus assumed impacts based on anecdotal information.

AVOID SELF-EVALUATION REPORTS

In two of the reports from fiscal year 2013 (Fulbright English Teaching Assistant Program, SportsUnited Programs), the data relied on

self-evaluation from those directing programs rather than the program participants themselves. The Fulbright English Teaching Assistant program asks English teaching assistants, rather than students, about teaching effectiveness and increasing understanding. Although there are other evaluation projects that focused on students/foreign participants, it’s difficult to justify such a primary focus on U.S. participants. Since the people involved in the programs are evaluating themselves and the contributions they made toward mutual understanding, it could lead to false conclusions regarding the effectiveness of the programs in achieving specific objectives. While language and travel constraints can make such self-evaluation data collection necessary, it is important to warn leadership about these limitations before evaluations commence and be forthright about it in long-term evaluation reports.

Recommendation: Reduce, as much as possible, the reliance on self-evaluation of program participants and flag the issue for leadership when long-term evaluations of programs are requested.

SUPPLY GREATER CONTEXT COUNTRY, REGIONAL AND GLOBAL TRENDS TO THE ACTIVITY BEING MEASURED

In-depth, qualitative data, which is one of the main methodologies in these studies, provide important sources of information. But the presentation of such data needs to be more contextual and thorough; otherwise it leaves the impression of being anecdotal. There is a lack of information about foreign participants and their communities, which may give insight into the varying worldviews that they hold. Often, the long-term evaluations also failed to capture the extent of program dynamics, such the differences in context—geographic, political, etc.—that could influence program outcomes. Time differences in terms of when participants took part in ECA programs also could skew the results. Multi-year studies tracking the same participants

over time could enhance understanding of the longevity of changes in knowledge, attitudes, and behavior.

Comparative studies, too, would contribute to the understanding of program effectiveness and the need for regional changes in program activities and/or implementation. Even though some of the survey respondents participated in the programs years ago, the aggregate level of analysis doesn't allow for any comparison between their more "long-term effect" and the responses from those who just completed the program, i.e., short-term impact. Comparative analysis, based on clearer data disaggregation, would allow for more relative impact analysis.

Recommendation: Place more emphasis on qualitative field research and comparative studies that would provide insights regarding the influence of context on program outcomes. Also incorporate multi-year studies to track long-term impact and changes in attitudes and behaviors in various contexts. This also requires more funding for measurement activities, especially those involving multiple-methods and time-series studies. Yet such data are also valuable for informing the design of future evaluation projects.

ENCOURAGE CONSTRUCTIVE CRITICISM THROUGH EVALUATION

The brief Performance Measurement Reports provided an overview of each program and highlighted findings based on survey data. Yet these reports, especially, provided a strikingly positive view of performance measures, which focused on self-reported changes in participants and included positive quotes from participants who filled out the surveys. Such an approach raises questions regarding both the validity and reliability of the data. Program highlights in each report varied, apparently according to program goals. There was no discussion of how each program supported the goals of the State Department and/or ECA more broadly.

Recommendation: Provide more objective data analysis in reports. Develop studies directed at detecting and understanding the reasons for both the formation of and shifts in attitudes and behavior among foreign publics toward the United States.

CLARIFY DESCRIPTIONS OF RESEARCH PROCESSES

The ECA evaluation division should strive to be more specific and systematic in describing the research processes it undertakes. For instance, in the Journalism and Media Exchange Programs evaluation, there was no discussion of the sample sizes and response rates for the global online surveys. Nor was there sufficient information about the focus groups and face-to-face interviews conducted. For instance, the response rates of the online global surveys for the International Visitor Leadership Program and the Murrow Exchange (as calculated based on total responses in tables and total participants in program overviews) were 3.4 percent and 13 percent, respectively. Such low response rates need explanations.

In the Murrow program evaluation, two different surveys were used: a global online survey of past participants, and a before-and-after analysis of the 2008 program. At various places within the report, these two different surveys were presented at the same time, making it difficult for the reader to clearly understand the findings and their contexts.

Recommendation: Further develop guidelines and principles that guide research and data reporting, including measurement that provides for learning and insights from evaluation to be used in improving and developing future programs. Aim to use qualitative research to inform quantitative research designs and not just illuminate the quantitative findings.

DISTINGUISH BETWEEN WHAT'S INFERRED VS. DIRECTLY ASSESSED OR OBSERVED

At times, data interpretation doesn't cohere with specific findings. For instance, in the Youth Exchange and Study Program report, participants' positive views of "the U.S. is a democracy that works well" and "the U.S. provides equal opportunity for all" decreased or remained at more or less the same level among several cohorts before and after their program experiences. But the narrative in the report doesn't adequately reflect that.

Also, most of the conclusions in these reports are inferred. One summary point in the English Language Specialist program states, "The majority of surveyed specialists believed that their work had impacted English language teaching in host countries at a national level." Another example, one specialist who had worked with a Ministry of Education to develop a "baseline exam" for all English teachers in the host country wrote, "I feel I have made a contribution to the teaching of English internationally." These statements need to be interpreted as claims, or they need to create detailed cases that may be able to substantiate these claims.

Recommendation: Provide objective measures of program impact on parti-

pants rather than inferences drawn from interviews with those involved in program planning and/or implementation.

CONCLUSION

As the oldest evaluation unit at the U.S. Department of State, we are encouraged by its moves to link evaluation more directly to U.S. foreign policy goals and provide more of a feedback loop for programs. Its budget constraints, representing just .25 percent (\$1.3 million) of the Educational and Cultural Exchange Programs budget, limit the amount of short and long-term evaluations undertaken and the depth of these reports. Increased funding for expanded field research, including cases studies, would provide valuable data on the impact of public diplomacy programming over the long term. Comparative studies also would enhance understanding of the impact of context on public diplomacy outcomes. Incorporating such methodologies in ECA measurement activities would help to develop a narrative showing how opinion has changed toward the U.S. and toward U.S. foreign policy over time among key publics. As identified in the opening of this report, the legal and organizational changes are also necessary to make measurement and evaluation a pervasive part of U.S. public diplomacy strategy and practice.

ENDNOTES

1. The Fulbright Hayes Act of 1961. <https://www2.ed.gov/about/offices/list/ope/iegps/fulbrighthaysact.pdf>
2. While the Journalism and Media Programs and Youth Exchange and Study Program included objectives and elements related to general policy goals, other programs lacked exploration of the linkages between the exchanges and policy priorities.

3. There is a similar problem with the goal of "cross cultural exchange and learning." For example, the SportsUnited Programs evaluation reported that "responding American envoys who traveled overseas communicated key American values, such as the importance of tolerance and inclusion, to the individuals and organizations who participated in their activities."

STATE DEPARTMENT: DIGITAL MEDIA EVALUATION IN THE INTERNATIONAL INFORMATION PROGRAMS BUREAU (IIP/ANALYTICS) AND PUBLIC AFFAIRS BUREAU (PA/ODE)

By Amelia Arsenault, Ph.D. AND Craig Hayden, Ph.D.

The Department of State manages approximately 1,000 social media accounts connected to diplomats, embassies, consulates, bureaus, offices, and programs around the world. Through these properties, it engages audiences in order to shape conversations on subjects pertinent to U.S. foreign policy objectives, and more broadly, to sustain relationships with foreign publics that build long-term trust and U.S. credibility.

There is currently no centralized methodology or office responsible for measuring and evaluating U.S. public diplomacy and public affairs activities conducted via social media. The International Information Programs (IIP) Bureau's Audience Research Measurement Office, now renamed the Analytics Office after an official reorganization, focuses its evaluations of digital activities mainly on campaigns and explores how content disperses among social media networks. The current emphasis of these evaluations is on analyzing the diffusion of messages, and less so on how social media content contributes to foreign policy strategy. Its team size fluctuates between three-five people.

The Public Affairs Bureau's Office of Digital Engagement (ODE) creates, manages and amplifies content for the State Department's flagship social media accounts, such as Twitter's @JohnKerry for the Secretary of State and @StateeDept for the agency; the department's YouTube channel; its Facebook page; and its Flickr page.

¹ One person on the team tracks and monitors the performance of these accounts to improve their effectiveness, which is measured by reach.² While analytics is the main mandate of IIP Analytics, this is not the case for ODE. The hundreds of other State Department social media accounts are maintained and tracked by social media co-

ordinators within various regional and functional bureaus, in addition to embassies.³

The following appraisal is based on written evaluations conducted with these two offices in 2013 and 2014 and contextual interviews with employees involved in evaluation. It outlines several mechanisms for rectifying the methodological, organizational, and structural limitations on evaluating U.S. public diplomacy outreach via social media.

State Department employees concerned with social media analytics are aware of many of the shortcomings outlined in this document and eager to implement reforms that enhance the utility of social media evaluations. IIP is currently in the midst of implementing several changes designed to address many of the programmatic shortcomings highlighted in this and previous reports.⁴ Most notably, it has begun to embed trained methodologists within teams tasked with public diplomacy campaigns, a welcome shift from commissioning evaluations after the fact. Moreover, under the IIP re-organization, the new Analytics Office will be located with the Campaigns and Regional Engagement Office of the bureau, a change that promises to better link research and evaluation to campaign design and implementation. This will provide evaluators with a seat at the table during the design phase, offer greater opportunities for campaign course correction, and promises to facilitate greater complementarity between campaign goals and evaluation techniques. These efforts are commendable and should be expanded.

While the Office of Digital Engagement (ODE) focuses mainly on reporting outputs with some brief descriptive analysis in its weekly and

monthly reports, we feel that given more expert capacity, they would be better positioned to also report the outcomes of how social media content is serving strategic foreign policy objectives. More systematic individual and holistic analysis of all U.S. social media properties and campaigns is needed. Few people outside of the State Department recognize or understand the bureaucratic divisions between what is considered public affairs and what is public diplomacy; thus, outsiders make no distinctions between which social media accounts are designed to fulfill a public diplomacy mandate and which are not. In order for evaluation to better inform how State Department social media properties serve programmatic and long-term foreign policy objectives related to public diplomacy, evaluations should ideally examine individual social media accounts and campaigns regardless of the originating office. They should also examine how the concert of State Department social media accounts complement or contradict one another.

There are several additional systemic hurdles to fostering an organizational culture that uses research and analytics effectively to assess tangible impact. Without significant structural reforms, officials' capacity to improve U.S. public diplomacy via social media is limited. The following pages offer several critiques of existing practices and highlight potential legal, budgetary, organizational, and methodological avenues for change.

BACKGROUND

As noted public diplomacy scholar Nicholas Cull reminds us, evaluating progress on long-term public diplomacy goals "can seem like a forester running out every morning to see how far his trees have grown overnight."⁵ While most public diplomacy activities are difficult to quantify, social media platforms—digital networks comprised of quantifiable and scalable human nodes around the world—on the surface might appear in comparison easy to monitor and evaluate. However, social media evaluation techniques writ large are still nascent because widespread

use of social media on a global scale is still relatively new. YouTube, Facebook, and Twitter only moved to global prominence in 2005, 2008, and 2009 respectively. The State Department launched its flagship accounts on YouTube, Facebook and Twitter in 2007 and 2008. Many embassies and bureaus began using social media in 2009 and 2010, but these social media tools were not commonly used until 2011. ECA developed the first social media platform based on the Ning service in 2007 and IIP launched its first Facebook pages in 2007 and its first Twitter feed in 2008. Moreover, every day, new social media start-ups targeting specific locales, languages, and proclivities emerge around the world. Each new platform evokes different challenges for public diplomats, whether they are legal (i.e., whether their policies conform to U.S. government guidelines), technical (i.e., whether the data be accessed and analyzed using existing methods), or social (i.e., how usage and interaction norms differ). Thus, public diplomacy social media evaluation efforts are not only emergent but must also adapt to an evolving technological environment, while confronting legal restrictions that inhibit government adaption of these technologies in the first place.⁶

Due to various bureaucratic limitations, not all U.S. public diplomacy social media initiatives are evaluated, and many evaluations are post-hoc and incomplete. Moreover, these evaluations, we understand, are not routinely read or consumed by program and social media implementers. As will be described in greater detail below, most of these methodological shortcomings are not directly attributable to State Department leadership or analytical staff, but rather to a combination of structural, cultural, and legal issues.

APPRAISAL OF CURRENT METHODOLOGY AND INTERPRETATION

Public Affairs' ODE and IIP's Analytics Office use different tools and operate according to different mandates. ODE conducts weekly and monthly descriptive analytics of the major social media properties maintained by the Office of Public

Affairs. The office mainly relies on commercially available tools (e.g., Google analytics, Bright Cove) to identify descriptive statistics on how many users follow, view and share public affairs content in social media spaces worldwide.

ODE focuses on the Department's primary social media accounts, while the majority of the Department's social media properties are maintained and tracked outside of Washington by embassy staff. At present, there are no systems in place to collect and track social media output at-large, though some solutions are being tested to manage the State Departments total social media presence.

The Analytics Office, in contrast, is responsible for the evaluation of social media properties and campaigns conducted by agents of the State Department falling under International Information Programs, including IIP employees headquartered in Washington, D.C. and U.S. embassies and diplomats around the world. Analytics conducts comparably more in-depth network analysis, exploring such issues as the centrality of particular social network accounts in conversations regarding a particular issue or theme and the resonance of particular social media outreach campaigns, as measured by proxy measures for "engagement" such as number of individuals retweeting or favoriting IIP initiated tweets. Until recently, the ODE had little formal coordination with the Office of Analytics; current efforts to improve contact between these units are encouraging.

While the specific quantitative metrics used in these analyses are methodologically sound, they are not necessarily sufficient to help understand the "impact" that public diplomacy has in advancing U.S. foreign policy. The majority of research into the efficacy of U.S. social media public diplomacy activities examines macro-level data trends. Due to legal, financial, and staffing limitations, evaluators are limited in the methods they can deploy and inhibited from examining micro-level audience data, a critical component of understanding the impact of U.S. social media

activities, due to the State Department's interpretation of the Privacy Act of 1974.

There are limitations to every method of analysis. Even when using the most robust research design, one cannot identify the precise contributions of social media campaigns to achieving foreign policy goals. However, there are key areas where State Department analytics and evaluation methodologies could be improved or expanded. Some of the suggestions below are not equally applicable across both offices. However, all social media activities consumed by or directed at foreign audiences are relevant to public diplomacy and should be evaluated with similar rigor, regardless of the originating office.

EXAMINE LEGAL RESTRICTIONS HAMPERING DIGITAL AUDIENCE RESEARCH

Because of legal restrictions on the government collection of individual-level data by the Privacy Act of 1974,⁸ public diplomacy evaluators can collect aggregate data about who is following, sharing, and viewing US government social media properties, but not by whom, why, and to what effect. Therefore, State Department social media evaluations provide little or no definition of the intended audience. Ideally, evaluations should provide a clear indication of (1) the intended audience, (2) the audience that was actually reached, and (3) a more nuanced picture of audience characteristics. This is not a new criticism, but one that has been highlighted by a series of previous reports.⁹

Recommendation: Social media activities should be conducted with a clearly articulated audience in mind. The ACPD, as noted in the overview document to this report, will further review the Privacy Act of 1974 with legal experts and make recommendations on whether or not it should restrict online audience research. In the meanwhile, evaluators should explore methods such as focus groups and interviews—which are not subject to the same legal restrictions—to better under-

stand how different audiences engage with DOS social media activities.

EXPAND THE METHODOLOGICAL TOOLKIT

Currently, social media evaluations are tailored to the methodological skill sets of existing personnel. While network analysis and descriptive statistics are useful, they provide only a partial depiction of “impact.” No complementary methods like interviews, experimental designs, content, and discourse analysis are currently used. Analysis is very often ad hoc, and done in a manner that inhibits systematic, longitudinal data gathering. This narrow methodological focus creates a system in which methods dictate the questions being asked rather than the reverse, which is the preferred research approach. There are limitations to the kind of claims that can be made about impact, if the data collected is only limited to the “message” itself.

Recommendation: Expand the range of methodologies used among the evaluation teams. While members of the analytical units have the methodological expertise, they could benefit from more trained field research staff, polling resources, and access to all available data spread across inter-agency units. They could also benefit from more direction from someone with business development expertise and/or social scientists with expertise in media effects, audience studies, and area or language-specific skill sets. Funding and staffing should also be increased to facilitate the expansion of baseline studies, audience research, and mixed methodologies.

DO NOT CONFLATE OUTPUTS WITH OUTCOMES

Descriptive statistics and even more advanced network analysis techniques do not establish causality. In other words, they provide little insight into whether what public diplomacy prac-

titioners do with social media tools influence the actions and attitudes of target populations. Analytics reports often conflate social media outputs (e.g., likes or retweets) with desired public diplomacy outcomes of positive sentiment or engagement. More social media activity (output) does not necessarily translate into increased engagement or positive sentiment (outcomes). Because no data is available (largely due to legal restrictions) about the characteristics or sentiments of individuals involved in social media exchanges with U.S. public diplomacy entities, there is no certainty about who is being reached and to what effect. Although evaluators acknowledged this limitation in person, written reports generally highlight increases in raw numbers and rarely comment on the extent to which these outputs contribute to desired public diplomacy outcomes. This can give a false impression of success.¹⁰

Recommendation: Evaluation reports should contextualize measures of outputs, define and qualify terms like “engagement,” and clearly delineate outputs from outcomes.

CONTINUE TO INTEGRATE EVALUATION AND POLICY UNITS WITHIN IIP

All the analytical units expressed a desire for a “seat at the table” when programs and social media activities are planned. Under the IIP reorganization, Analytics will be more closely associated with their Campaigns Office, which should help to better facilitate integration between evaluation and policy. This is a movement that should be encouraged and expanded, as evaluations are only useful if they are (1) used to inform subsequent activities and (2) designed in such a way that they clearly measure the goals of the project or initiative. More thorough analysis can be conducted if evaluators are involved at every stage of analysis, especially in creating baseline data that can later help interpret successes or failures. It also allows researchers to design monitoring systems for programs while they are being conducted rather than after, and to recom-

mend minor tweaks to implementation that may make data collection easier.

Recommendation: Wherever possible, social media evaluation plans should be designed *before* the initiation of new programs so that stated goals and metrics of evaluation can be better aligned. Program design and long-term strategic planning should also draw upon the knowledge and input of previous evaluations and evaluation team members. Evaluation teams should also be tasked with providing project implementers with evaluations of prior complementary efforts. Conversely, project implementers should be tasked with consulting evaluation teams *before* designing outreach campaigns.

BUILD CONSENSUS EARLY ON WHAT IS BEING MEASURED AND WHY

Interpreting program success is difficult without clearly articulated goals. There needs to be better definition of foreign policy goals and the objectives of social media campaigns and messaging. Even if detailed audience data were available, administrators and implementers rarely provide evaluators with clearly defined desired outcomes and goals against which social media activities should be evaluated.¹¹ In IIP, based on our interviews with officials and from the reports provided from fiscal year 2013, evaluations were almost exclusively initiated and designed after outreach campaigns were completed, and even then the goals of the program were rarely communicated to the evaluators. Evaluations also pay little attention to articulating how social media efforts contribute to achieving long-term public diplomacy goals.

Recommendation: Clearly align program goals and outcomes with metrics, tying public diplomacy activities to foreign policy objectives. Evaluations need to address how social media activities fulfill desired short-term campaign outcomes as well as how they achieve (or don't

achieve) longer-term public diplomacy goals. More baseline studies and robust longitudinal investigations that incorporate methods like panel studies would help to shed light on the contribution of social media efforts to long-term public diplomacy efforts.

SUPPLY GREATER CONTEXT ABOUT DEPARTMENTAL, COUNTRY, REGIONAL AND GLOBAL TRENDS TO THE ACTIVITY BEING MEASURED

Social media activities do not take place in a vacuum. Appraisals rarely, if ever, contextualize the performance of State Department activities within larger online and offline trends. They generally focus on very small network communities and present traction within these communities as a generalized success. Overlooking extraneous technical, social, or political factors outside the boundaries of the campaign can give a narrow interpretation of public diplomacy campaigns' successes or failures. Social media platforms, moreover, are embedded within a network of old and new media networks. There is an unrealized opportunity to examine how social media parallel, reinforce, or contradict narratives circulating within "old media" and attitudes and beliefs held by participants in offline programs, such as exchanges and visitor programs. Moreover, more systematic individual and holistic analysis of all US social media properties and campaigns is needed.

Recommendation: Employ methodologies that account for the interaction of interpersonal, cultural, and mass media effects in order to demonstrate how public diplomacy campaigns create impact and how such efforts demonstratively contribute to short-, medium-, or long-term policy. Complement evaluation and analysis of individual accounts and campaigns with evaluations of social media activities related to public diplomacy and public affairs Department-wide.

COMPARE U.S. ACTIVITIES WITH THOSE OF INTERNATIONAL ACTORS

ODE and IIP Analytics rarely include comparative data in their evaluations. Whether one is evaluating the success of a hashtag or a particular theme, it is difficult to measure success if there is no point(s) of comparison. Contrasting U.S. efforts with comparable efforts from other international actors can provide a constructive basis for analysis and course correction. There is also the unrealized opportunity to compare current social media campaigns with past campaign performance. Compiling a longitudinal basis for analysis and comparison may be crucial to developing both audience understanding and a more nuanced repertoire of messaging tactics under specific circumstances.

Recommendation: Incorporate benchmarks into evaluations that compare U.S. social media activities with those of analogous actors (e.g. allied countries involved in similar campaigns and outreach). State may also consider formal evaluation sharing agreements with allied countries or other U.S. government actors such as the U.S Agency for International Development.

ENCOURAGE CONSTRUCTIVE CRITICISM THROUGH EVALUATION

Evaluations provided by IIP and ODE often highlight successes and omit or downplay shortcomings, although ODE does indicate in monthly reports trailing data. As is the case with almost all bureaucracies, suggestions of limited or negative outcomes may inhibit future funding and administrative support. An expectation of success creates a climate that inhibits realistic evaluations. In order to provide meaningful criticism as “course corrections” for campaigns, measurement and evaluation units need the authority to provide candid and thorough guidance, and program implementers and management staff should be encouraged to view analytical products as constructive (rather than punitive).

Recommendation: Reward and encourage honest and balanced assessments. Officials should be empowered to write evaluations in a more balanced manner that highlights the successes and *failures* of particular campaigns and activities. Bureaucratic reorganization that integrates policy and analytical units, such as the proposed movement of a re-branded Analytics office grouped with the Campaigns and Regional Engagement office, is an important first step in fostering appreciation for honest evaluations.

INCREASE DATA SHARING ACROSS STATE DEPARTMENT

In the past, because analytical units are embedded in the chain of command of particular bureaus (e.g. IIP or PA), convoluted reporting lines and bureaucratic jurisdictional disputes hampered the widespread circulation and usage of social media evaluations.¹² The average State Department employee involved in social media activities is not a trained consumer of social media analytics and rarely consults previous evaluations when planning new activities. While there are no clearly articulated procedures or incentives for sharing social media data and analytics reports, this appears to be improving. We’re encouraged by the increased communication and improved cooperation within and between the Public Diplomacy and Public Affairs cone’s bureaus compared to yesteryear.

Recommendation: A team should be tasked with aggregating and sharing evaluations across departments engaged in international outreach activities to facilitate systematic and sustainable procedures for pre-and post-communication analytical efforts. A Research Director in the public diplomacy and public affairs front office, as noted in the report’s overview, could help press for systematic collaboration. In addition, for the department’s non-flagship social media accounts, access to a common content

management system (e.g. Hootsuite is currently being tested), could enable better coordination of social media content across DOS accounts and support more longitudinal studies. Capacity and authority should be expanded so that all DOS social media properties relevant to PD are subject to the same rigorous evaluation standards in order to better inform DOS-wide social media activities. Consumers of analytical reports need to be trained to read and interpret findings and encouraged to seek out evaluations of previous or complementary social media activities.

CONCLUSION

Despite the often ad hoc nature of the analysis conducted, there is clear evidence that State Department employees are increasingly appreciative of the value of analytics and evaluation

for understanding and improving public diplomacy activities conducted via social media. The analytical officials dedicated to measuring the impact of U.S. social media output for public diplomacy are enthusiastic and dedicated. They have developed innovative workarounds to data access issues and new social media content management and data sharing solutions. However, much of their analysis is constrained by: (1) an organizational culture that inhibits aligning metrics with goals and outcomes; (2) the absence of an institutionalized means of circulating and sharing data and evaluations; (3) barriers to accessing individual-level data; and (4) financial, legal, and staffing constraints on expanding their methodological toolkit. Given these limitations, their work is commendable. However, legal and organizational changes are needed to bring the insights of these measurement and evaluation personnel into all stages of the public diplomacy process in order to better inform future public diplomacy campaigns.

ENDNOTES

1. These include: DipNote (<http://blogs.state.gov>), Facebook (<https://www.facebook.com/usdos>), Twitter (@StateDept, @USAAbilAraby, @USAAdarFarsi, @USAEnEspanol, @USAEnFrancais, @USAPoRuskki, @USAIndiMein, @USA_Zhongwen, @ABDTurk, @USAemPortugues, @USAUrdu), YouTube (<http://www.youtube.com/user/statevideo>), Google+ (<http://www.youtube.com/user/statevideo>), Flickr (<https://www.flickr.com/photos/statedphotos>), Tumbler (<http://statedept.tumblr.com>) and streaming video and other engagement delivered via the State Department website.

2. In both the case of ODE and IIP/Analytics, the leadership chain in both offices has changed significantly since the time of the review period. As a result, in most cases the interviews did not take place with the leaders in the 2013 calendar year who would have been familiar with the evaluation processes at the time.

3. In an effort to coordinate the analytics of the hundreds of State Department social media accounts that are outside the scope of ODE, the Policy, Planning and Resources Office in the public diplomacy cone is in the pilot phase of rolling out a shared social media delivery platform via Hootsuite, which will better offer social media coordinators in regional bureaus access to real time data as well as the ability to better coordinate and share content across the diverse array of bureaus and diplomatic posts involved in social media outreach. The Office of Digital Engagement has decided not to use Hootsuite to measure the analytics for the flagship State Department social media accounts.

4. The most significant reports that have inspired reforms in social media evaluation practices are a 2013 report by the United States Department of State and the Broadcasting Board Of Governors Office Of Inspector General “Inspection of the Bureau of International Information Programs” and the 2012 the internal Social Media Working Group Report,

which drew upon the input of 60 key stakeholders from across the State Department’s public diplomacy and public affairs operations.

5. Cull, Nicholas J. “Public Diplomacy: Taxonomies and Histories.” *The ANNALS of the American Academy of Political and Social Science* 616, no. 1 (March 1, 2008): 44.

6. For more on this, see: Bertot, John Carlo, Paul T. Jaeger, and Derek Hansen. “The Impact of Polices on Government Social Media Usage: Issues, Challenges, and Recommendations.” *Government Information Quarterly* 29, no. 1 (January 2012): 30–40. doi:10.1016/j.giq.2011.04.004.

7. The State OIG made a similar observation in 2013, “Facebook analytic tools can measure engagement by counting the number of people who click on a link, “like” a posting, comment on it, or share it with their friends. However, these measures do not evaluate the usefulness of the engagement because many people post simple remarks, like “so nice pic,” or comments on unrelated topics (2013: 22).”

8. A July 2013 State OIG report noted, “The bureau [IIP] could reduce spending and increase strategic impact by focusing its advertising not on raising overall fan numbers or general engagement statistics but on accomplishing specific public diplomacy goals. The Bureau of International Information Programs should adopt a social media strategy that clarifies the primary goals and public diplomacy priorities of its social media sites” (GAO 2013: 22). The majority of the evaluations we have evaluated were conducted before the release of this report.

9. This was highlighted in the 2013 OIG Report: “There is limited communication and no regular meetings between PA and IIP to discuss what each bureau is doing in this area, leading to duplication of effort. (OIG 2013: 23).”

BROADCASTING BOARD OF GOVERNORS: RESEARCH AND EVALUATION APPRAISAL

By Shawn Powers, Ph.D., Matthew Baum, Ph.D., and Erik Nisbet, Ph.D.

As the U.S. government's international broadcasting agency, the Broadcasting Board of Governors (BBG) produces audience research that is a valued resource not only to American broadcasting practitioners, but also to foreign international broadcasters, including the BBC and Deutsche Welle. Collectively, this diverse set of research offers an impressive array of data that should be of help in understanding the news consumption habits of global audiences. They also do so with extremely tight funds: In fiscal year 2015, the proposed budget is \$4.7 million, or .7 percent of the overall \$721.26 million budget. This is far below the 5 percent industry standard.

We reviewed more than 60 reports from Gallup, an independent, for-profit research organization, completed during fiscal year 2013 (October 2012 through September 2013) and a handful of documents from 2014.^a The reports were focused on Voice of America (VOA), Office of Cuba Broadcasting (OCB), Radio Free Europe/Radio Liberty (RFE/RL), Radio Free Asia (RFA), and the Middle East Broadcasting Network (MBN). We welcome the reports' multifaceted methodological approach. Specifically, they are most effective when they combine basic audience ratings with in-depth interviews and focus groups. These three methods, when employed in tandem, provide a richer, more holistic context for understanding U.S. international broadcasting's successes and challenges than any one or two of them can independently provide.

The BBG staff and board are making two commendable moves toward the generation of audience research and impact assessment. We support BBG's move towards an Impact Framework for measuring their activities' effectiveness. Working closely with Gallup, the new model attempts to examine each of the micro (individual), meso (society/community) and macro

(government/institutional) levels over short-term, mid-term, and long-term timescales to measure how programming makes a "difference in the lives of our audiences in ways that correspond to" the BBG's mission of informing, engaging, and connecting people around the world in support of freedom and democracy.¹ Though the BBG Impact Framework could be improved, we believe it has substantial promise of improving BBG evaluation in the long-term (see Appendix I for the BBG Impact Framework and a full review of it).

We also strongly support the reorganization of existing research capabilities and the merger of elements of the Office of Performance Review and the Office of Strategy and Development in the International Broadcasting Bureau (IBB), both established in 2012, to create the IBB Office of Research and Assessment (ORA). The move should enable the BBG to take full advantage of the rapid evolution of digital data analytic tools, techniques, as well as the visualization of that information, and seamlessly merge these techniques with more traditional analysis of survey and acquired data. The ORA will substantially improve the strategic focus and efficiency of research efforts by overseeing the entire research and digital infrastructure, including the introduction of new information intelligence products and tools, for all of BBG.

While BBG is implementing the Impact Framework and has adopted a revised survey instrument to specifically measure research and non-research indicators, there are valuable lessons to learn from the 2013 and 2014 reports reviewed. The surveys and focus group interviews emphasize gathering information concerning where people get their news, who gets news from different sources, at what times of day, and other broad questions regarding media habits and behavior. Yet it is unclear in most instances

what to take away from these data and what the strategic focus for the reports are. That is, who are the target audiences? Why? What are the intended effects of consuming U.S. broadcasting? What are the specific strategies for reaching those intended audiences?

We appreciate that the Gallup reports are not the only research products the BBG produces or commissions, and that there are many research functions to respond to the various requirements of the Government Performance and Results Act (GPRA), to Congress, and to the Government Accountability Office (GAO). Yet absent a clear *ex ante* strategy for pursuing this research with Gallup, it appears somewhat haphazard *ex post*. This gap may also hamper the implementation of the new Impact Framework and survey instrument, and continue to create a disconnect between their conceptual and methodological dimensions (see Appendix I).

BACKGROUND

The mission of the BBG is to “inform, engage and connect people around the world in support of freedom and democracy.”² This mission is distinct from the State Department’s public diplomacy activities in that its primary objective is not to persuade attitudes regarding the United States or its policies. It is, however, strategically aligned with broader U.S. foreign policy goals. For example, BBG’s 2013 Annual Report lists several targeted goals achieved through its networks, including: (1) exemplifying free press to help foster and sustain free, democratic societies; (2) serving as a model of free press; (3) combatting violent extremism; (4) fostering greater understanding and engaging audiences.

BBG’s research aims to measure the efficacy of its programming in achieving some or all of these objectives in the target countries within which it operates. Specifically, it measures effectiveness in terms of audience size, program quality and reliability, and whether or not programming increases the audience’s understanding of current events and American society and policies.

The BBG contracts with Gallup to conduct qualitative and quantitative audience and market research. Gallup focuses primarily on audience research, employing a mix of surveys, focus groups, in-depth interviews and audience panels. BBG defines audience size as the number of adults (15+) who “listened, viewed or used online material last week” based on random sampling in the target area. (This is the agreed-upon standard used by most international broadcasters.)³

APPRAISAL OF CURRENT METHODOLOGY AND INTERPRETATION

Overall, the research reviewed was rigorous and scientifically sound. Audience samples are generally large and allow for a relatively small margin of error. Gallup does an effective job of ensuring that the methods deployed follow established, contemporary scientific rigor. That said, we feel there is room for improvement for future studies. We identified several concerns regarding the research design, methodologies and instruments used that, if addressed with the proper resources, would increase the value of the research as the BBG moves more toward understanding its worldwide impact.

INTEGRATE RESEARCH METHODOLOGIES

There was a disconnect between the large, general survey studies and the in-depth interview and focus group studies. Put another way, there could be a tighter relationship between the large-N surveys and small-N focus groups. At the outset, it is hard to place too much confidence in standalone focus group-based conclusions, as they are, by definition, anecdotal. However, they can be valuable for fleshing out insights derived from larger-N studies or controlled experiments. In one instance, a Gallup report on Cuba Broadcasting shows results from 191 focus group interviews and a complementary 295 in-person, in-depth quantitative interviews. This approach may more likely yield generalizable results than interviews with 15-20 individuals. Of course, this all depends on

having the right participants (samples based on clear conceptual and methodological criteria representing specific audience characteristics, behaviors, or groups of interest), and, in most instances, it was unclear upon what criteria (strategically speaking) the samples were selected.

Recommendation: Develop a methodologically integrated approach that combines large-N surveys, medium-N in-depth interviews, and small-N focus groups in a conceptually coherent sequence. Begin with large-N surveys to identify salient issues for further study. Then turn to medium-N in-depth interviews to probe the key issues raised by the survey analysis. Finally, employ carefully selected (driven by the issues being investigated) small-N focus group interviews to flesh out the insights that emerge from the prior levels of analysis.⁴

CONNECT REPORT DESIGNS MORE WITH STRATEGIC OBJECTIVES

Reports using in-depth interviews to investigate the habits and impressions of particular audiences, while important, seemed disconnected from U.S. international broadcasting's strategic objectives. These reports focus on specific reactions to programming or new platforms, but fail to connect these findings to broader metrics measuring the impact of U.S. international broadcasting (e.g., freedom and democracy).⁵ While there is certainly value to this research, its connection to BBG mission needs to be made more clear (e.g. why target a particular audience in a particular country?).

Studies requiring participants to access BBG content online or via DVD prior to an interview need to be careful to account for how the introduction of particular programming may shape the reported findings. For example, studies required participants to listen to different radio programs, read news stories via their mobile phone, "like" Facebook pages and monitor updates, and/or watch web interviews while being

interviewed. This methodology may substantially bias research outcomes and estimated impacts of such content relative to real-world, naturalistic consumption, especially among participants unlikely to select such programming in the first place. While there is value to exposing participants in order to evaluate specific programming or content, the value of any findings are limited and should not be used to make broader claims about the success of failure of a broadcaster or program.^b

Recommendation: Use the large amount of data currently gathered by BBG more. We recommend controlled experiments (both field- and survey-experiments) designed to more rigorously test some of the conclusions drawn from this data, as well as to validate the utility of survey questions. Panel research designs that include both BBG weekly audience members and non-members (as a control-group), and quantitatively assess knowledge acquisition and attitude formation over time, should be implemented. Current planned panel and survey research does not meet this standard, though, RFE/RL is moving in the right direction by including participants not accessing BBG programming in its 2014 Internet panels.

In addition, questions querying participants' recall need to be reformulated where possible. People are more likely to remember specific programs they watch more than channels or outlets. Audiences are also better at assessing their most recent media consumption behavior than their longer-term consumption. We recommend evaluating whether additional questions focused on specific programs, and fewer questions on networks or outlets, geared toward emphasizing near-term recall (e.g., "yesterday" rather than "the past week"), might yield more accurate estimates from respondents for some markets.⁶ This, in turn, may allow richer data collection about the impact of BBG

programming, as well as provide an opportunity to assess the validity of current methodologies.

RECONSIDER DATA COLLECTION IN REPRESSIVE AREAS

In authoritarian, non-democratic survey contexts, posing explicit questions to respondents about sensitive topics such as domestic governance, foreign policy, or the United States could introduce substantial bias into survey results. For instance, in the March 2012 BBG Iran Report, respondents were asked “As you look towards the future, which system do you think is best suited to Iran?” Surveys that ask similar questions in other authoritarian contexts often use a more subtle approach, asking general preferences for different forms of governance, rather than specific preference about the respondent’s own country, as means to limit response bias and increase comfort answering the question. Peer-reviewed scholarship on this question suggests that audiences in deeply controlled, authoritarian regimes may be comfortable answering survey items about media consumption and attitudes (where the potential for sanction is low), but at the same time be uncomfortable answering survey items on the same instrument regarding political or foreign policy topics (where probability of sanction is higher).⁷

Recommendation: Reconsider how to best access sensitive information from audiences in repressive, controlled-information environments (e.g., Iran). For instance, item analysis examining rates of refusal or “don’t know” responses and survey wording experiments aimed at identifying and quantifying these biases in such countries could provide greater confidence in the accuracy of BBG survey results on potentially sensitive topics.

LESSEN RELIANCE ON SELF-REPORTED DATA

An important metric studied—self-reports of “learning” from BBG programming—could use

improvement. Put simply, people stating they learned something is not the same thing as having done so. Within academic scholarship and peer-reviewed research on media effects these survey items would not be considered valid and accurate means of measuring such media influence. A more valid and accurate means of measurement would be to ask three to four knowledge (with “knowledge” defined as factually and contextually accurate information) questions on each topic to assess accuracy of audience beliefs (for examples of localized knowledge questions, see the World Values Survey). The existing self-report measures are highly abstract and ambiguous as learning can take on different forms. For example, one can “learn” to appreciate the United States, but research has also demonstrated that some audiences have “learned” from American media to strongly dislike the United States.⁸ In addition, audiences may easily “learn” inaccurate beliefs or misinformation from the media.⁹ To measure learning, a key metric in the Impact Framework, we would suggest determining precisely what they learned, the accuracy of the garnered beliefs, how much learning knowledge retention takes place, and from which programs. In their current iteration, these and related questions are too broad to be very useful (see Appendix I). At minimum, inclusion of select scales on a trial basis would allow for reliability checks. That is, the results from direct self-reports could be compared against results from scales based on sets of questions aimed at capturing the underlying construct (attitude or behavior) of interest.

Recommendation: Expand the core BBG country questionnaire to include audience attitudes and assessments of democracy, press freedom, internet freedom, corruption, accountability, and self-efficacy in order to understand how the BBG influences core drivers of democracy, media sustainability, and accountability within its audience over time, among countries, regionally, and globally.

EXPAND BEYOND “AUDIENCE REACH” METRICS

Reach—the number of people who accessed BBG content in the past week—is the first thing that outside observers ask about, yet it has limited value. It offers little information about how much programming people actually consume, what they remember, what they think about the programs they watched, or how what they consume, remember, or think influences attitudes toward freedom of expression or democracy. The weekly reach metric (WRM) can also be misleading. For instance, Voice of America (VOA) Radio’s weekly reach in Bangladesh is 2.1 percent, while the BBC’s is 6.4 percent. While both percentages are small in an absolute sense, the corresponding figure for the market leader (Bangladesh Beter) is only 9.1 percent. One might conclude either that all is well, as VOA is doing nearly as well as the market leader, or that no one, including the market leader, appears to be reaching a mass audience in Bangladesh. The point is that while the weekly reach metric is a helpful starting point for analyzing the potential significance of BBG programming, it is insufficient for measuring “impact.” We thus suggest de-emphasizing the weekly reach metric, and instead placing greater emphasis on the significance and meaning of reach in particular environments.

Recommendation: Knowing about audiences that already tune into BBG programming is important, but can lack insight into why others are not tuning into BBG programming or why they don’t find it compelling or credible. Analysis of non-listeners needs to be institutionalized in the evaluation process in order to extend the reach and potential impact of BBG content.^c The Impact Framework includes other indices and factors, and we encourage their inclusion (see Appendix I).

MAINTAIN CONSISTENCY IN RESEARCH QUESTIONS, SURVEY MODALITY, AND RECOMMENDATIONS FOR LONG-TERM TREND ANALYSIS

A lack of consistency in reported research questions across the reports reviewed makes it hard to assess which questions are not being asked, or, more importantly, which of those should be asked. Put another way, the reports highlight particular findings from each country or sample and are tailored to tell a particular story about a country or target audience. It is unclear whether Gallup has based this tailoring on some ex ante research strategy or purely on ex post reviews of results. While this approach may have some benefits, it results in a lack of consistently comparable sections and questions vis-à-vis other countries, making it difficult to make meta, comparative assessments, or even to compare across two reports. For example, several of the reports tied in audience impressions towards freedom of expression, while others did not. Given the centrality of promoting a free press to the BBG’s mission, it seems as though this would be an important area of inquiry and crucial to measuring the impact of BBG’s programming.

We’re also concerned about changes in question wording and survey modality over time and what that might mean for talking about trends in a meaningful way. Gallup is forthright about the problems, and one of the reports we reviewed presented some evidence that despite changes in wording and modality the reported trend appeared meaningful. But in most cases they did not offer any persuasive evidence that these over-time comparisons are apt. This raises the question of what we can, or should, take away from these trends when the question wording, survey modality, sample demographics, or other things significantly vary across surveys? If the aforementioned flaws fundamentally compromise the trend analyses, then is reporting the trends misleading and does it produce analysis and conclusions that are themselves misleading?

Each of the reports prepared by Gallup included recommendations, but the basis for the recommendations oftentimes seemed anecdotal—rephrasing something that a focus group participant or two said—or at most loosely based on the larger-N data. It was unclear why certain recommendations were highlighted at the expense of others, or why we should believe the recommendations as plausible. There were several examples where the reports’ central findings or recommendations seemed to contradict data that was later described in detail.¹⁰

Recommendation: We think it is potentially unwise to describe pairs of discrete surveys as “trends” if differences between them make such conclusions unreliable.¹⁰ Another approach would be panel designs, which BBG and Gallup are implementing in 2014 on a trial basis. We support BBG efforts in this direction, and encourage piloting something akin to Nielsen-type monitoring of actual media usage for a sample population; emphasizing television and/or online media consumption would also be helpful. This could be married with snap polls and interviews to create a much more rigorous estimate of what people are watching, why, and what they are taking away from it. Where feasible, we encourage snap polls to be administered in real time, while people are watching or consuming Internet sites or even television content.¹¹

SUPPLY GREATER CONTEXT OF REGIONAL AND GLOBAL TRENDS AND BENCHMARK ACTIVITIES AGAINST THOSE OF INTERNATIONAL ACTORS

A conspicuous weakness that emerged when reviewing the individual reports and analyses supplied for review was a lack of regional or global integration of the surveys and data. There was no comparative report or analyses that benchmarked countries against each other, provided comparative context, examined regional

or global trends in BBG broadcasting, or tried to integrate contextual, country-level factors with survey data on a quantitative, analytical basis. Though we acknowledge the goal of BBG is to be “hyper-local” and provide value-added broadcasting to national audiences, from a strategic and analytical perspective, research should be regularly analyzing regional and global trends with such integration and analysis being a standard component of their reporting. This would allow more sophisticated quantitative cross-national, multi-level analysis of survey data¹² that statistically disaggregates contextual factors (e.g., macro/national level variables such as economic development, media infrastructure, ethnic fragmentation, democratic governance, press freedom, etc.) from individual-level factors (e.g., age, education, gender, media behaviors, economic status) in terms of predicting BBG impact outcomes (e.g., reach, learning, etc.).

Recommendation: Employ more advanced statistical methods for analyzing cross-national survey data, such as hierarchical linear modeling (multi-level analysis) of aggregated cross-national survey data to identify and measure global and regional predictors of BBG impact.

DEEPEN ANALYSIS OF SURVEY RESULTS

There is relatively limited “deep” analysis of the survey results, and what analysis there is seems at times to extrapolate beyond what the data presented actually show into varying degrees of speculation. Given its experience and expertise in this area, additional independent analysis by Gallup would increase the utility of the data collected.

Recommendation: Opening up parts of the BBG/Gallup data to a trusted community of academics and stakeholders for peer review on a routine, annual basis would offer additional analysis of what data mean, as well as an additional validity check on the suggestions that emerge in the reports. This could coincide with

expert workshops with academics, key stakeholders, and pollsters aimed at developing a set of achievable goals for this research. Key questions might include: What is the operational purpose of this broadcasting? What are we trying to accomplish from a strategic perspective? What concepts can be operationalized and how? What concepts cannot and why? The overview document in this report recommends an ACPD sub-committee that could help serve this purpose.

CONCLUSION

Given the changing nature of media consumption and production around the world, and the diversity of environments the United States broadcasts to, the BBG is leading a re-evaluation of the metrics deployed in measuring success, which is encouraging. This transition is an important opportunity to ensure future research designs and methods are sufficient to show the impact of BBG programming, as well as offer insight as to how programming could be improved.

The lack of a consistently applied strategic framework guiding fiscal year 2013 research likely reflects the challenge of operating in countries where audiences have different needs and expectations. One way forward would be to create categories of countries based on levels of development and informational needs. For example, four possible categories reflecting different missions could include: strategically significant, priority countries with populations at risk

of becoming radicalized; closed societies where media are largely controlled by the state; transitioning countries without sufficient competition and/or protections for freedom of expression; and failed, near-failed or weak states lacking sufficient authority for basic governance.

From these (or other) categories, specific research questions and scales could be explored through surveys, panel studies, focus groups, in-depth interviews and experiments. This would still require a core survey instrument with questions asked across countries, but would also include specific category-dependent modules that go into greater depth and measure the effectiveness of achieving specific goals, including combatting extremism, promoting freedom of expression, media pluralism, rule of law, and democratic governance.

We urge research products be used to help set strategic goals for the multiple language services within VOA and the surrogates, and that the appropriate directors communicate with researchers on what those strategic goals are. In addition, in order to properly implement new research techniques and the new Impact Framework, as well as the recommendations outlined herein, we strongly endorse an increase in the BBG research budget. In fiscal year 2015, research effort will account for less than one percent of the BBG's budget, well below industry standards (5 percent of the operating budget). Additional resources are required for BBG to effectively achieve its crucial mission in today's highly competitive and fast-changing media environment.

PRELIMINARY ASSESSMENT OF THE BBG IMPACT FRAMEWORK

The BBG provided this evaluation team four documents for review regarding the new BBG Impact Framework: a PowerPoint presentation providing a general overview of the Impact Framework and measurement, a survey questionnaire representing the most recent version of the standard BBG country survey designed to align with the Impact Framework, an excel worksheet providing a list of variables measured within the Impact Framework and their corresponding survey items on the BBG survey questionnaire, and a document providing a methodological overview of an online BBG panel impact study to be conducted in selected countries. Our assessment is based on these documents.

The BBG Impact Framework and the BBG Strategic Plan 2012-2016 define impact as “ultimately about whether we make a difference in the lives of our audiences in ways that correspond to the larger aims of our efforts.” The BBG Impact Framework has three goals over three different timescales: (1) to inform audiences in the short term (2) to engage audiences in the mid-term, and (3) to promote change in the long-term. BBG’s envisioned “outputs” are to reach key audiences and develop local media with “outcomes” being an informed public, an established sustainable local media, and government accountability.

Measurement within this Impact Framework has been conceptualized and operationalized for each goal. For example, whether the BBG informs audiences is based on (1) reaching the target audience as measured by weekly reach among the general population, vernacular speakers, and target segments and (2) providing value as measured by audience perceptions of exceptional information, trust, and level of interest.

Overall, we laud the BBG Impact Framework as a major strategic step forward in specifical-

ly outlining how the BBG conceptualizes and assesses impact across key dimensions and different timescales. Based on the documentation provided, we discovered some weaknesses to it in terms of operationalization and measurement, especially in when measuring change among audiences in the long-term. Thus, we aim to provide some constructive feedback with the goal of furthering the new BBG Impact Framework, which we believe has substantial promise in the long-term.

Specifically we recommend:

1. INFORMING AUDIENCES

There appears to be no global standard for what defines a “target audience” when BBG measures weekly reach within “targeted” audiences, as each country and/or broadcaster has its own definition for “targeted.” We suggest adding an additional indicator to the BBG Impact Framework that assesses weekly reach among what are typically called “opinion-leaders” or “influencers” in each country as a comparative, global standard of impact. Previous research suggests that measures of interpersonal discussion and/or perceived personal strength are the best means of identifying opinion leaders on cross-national surveys. Such an audience segment is most likely to share and pass along information from BBG broadcasts to other members of their social networks who may not directly access BBG programming, as well as recommend BBG to others.

This “two-step” flow is critical for information diffusion and dissemination and should be a standard component of BBG impact evaluation. Though the BBG measures whether audiences share news they have heard, seen, or read from the BBG, this measure does not capture the reach specifically within this target

segment. Previous scholarship has demonstrated that frequency of interpersonal discussion is one of the most valid cross-cultural indicators of personal influence and opinion leadership.¹³ Therefore, weekly reach within those audience members that share news “daily or most days per week” or “at least once a week” (survey item G4) should be standard indicators of targeted reach that the BBG includes in its impact framework.

2. AUDIENCE ENGAGEMENT

Likewise, audience engagement should also be measured among “opinion-leaders” specifically, as well as the general population, as part of the BBG Impact Framework. These indicators may be constructed by calculating the percentage of weekly audience that 1) “shared something they hear/read/see on broadcaster weekly” and 2) were “likely to recommend the broadcaster to others” among weekly audiences who have shared news “daily or most days per week” or “at least once a week” in general (survey item G4). This indicator should provide some insight into how much the BBG is impacting interpersonal discourse within the most influential segment of its weekly audience.

3. PROMOTING CHANGE

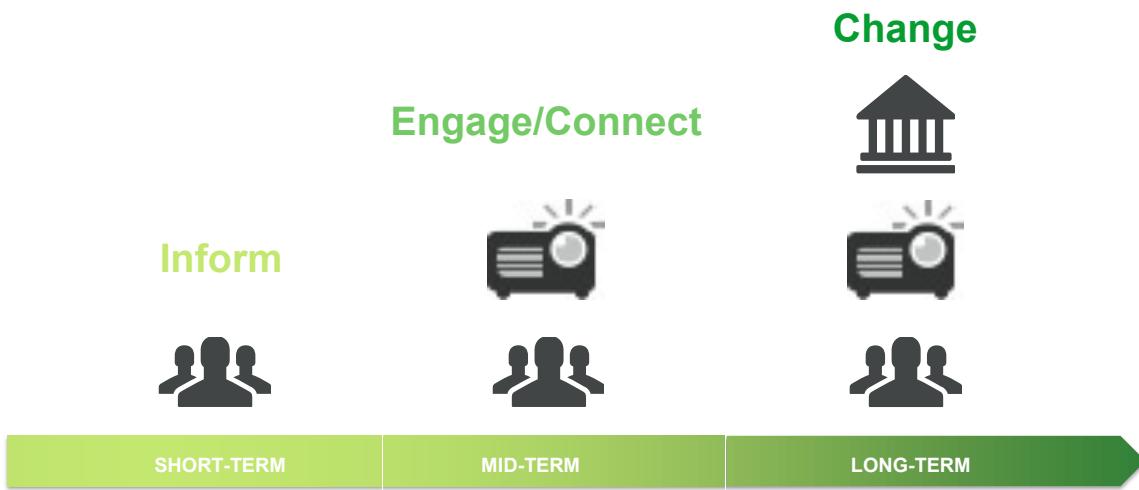
In our view, the biggest disconnect between conceptualization and effective measurement occurs within the BBG attempt to evaluate its influence on audiences and contextual change (e.g., democratization, sustainable free media) over the long-term. For example, as mentioned elsewhere in our report, the BBG assesses whether it has increased understanding of current events, American society, and U.S. foreign policy by single item measures asking respondents to self-assess their own understanding and the BBG influence on it. Within academic scholarship and peer-reviewed research on media effects these survey items would not

be considered valid and accurate means of measuring such influence. A more valid and accurate means of measurement would be to ask three to four factual knowledge questions on each topic to assess accuracy of audience beliefs.

The BBG Impact Framework also attempts to assess whether BBG programming increases audience understanding of democracy. Again, such self-reports of media influence are not generally considered valid or accurate. Moreover, we would assert that increasing understanding of democracy is not as important as increasing citizen *demand* for democracy among BBG audiences as BBG promotes “bottom-up” democratization.¹⁴ In addition, asking BBG weekly audiences how much media freedom they *perceive* they have in their country and how much they value or “demand” press freedom are also important indicators of free media sustainability as such attitudes are important for both promoting change and sustaining media freedom in the long-term.¹⁵

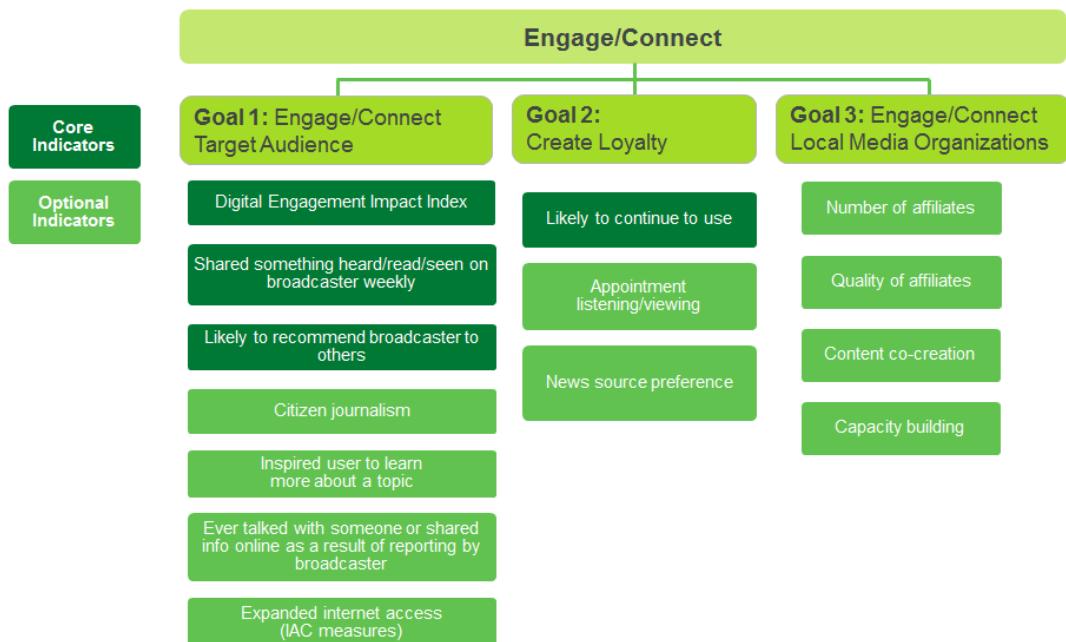
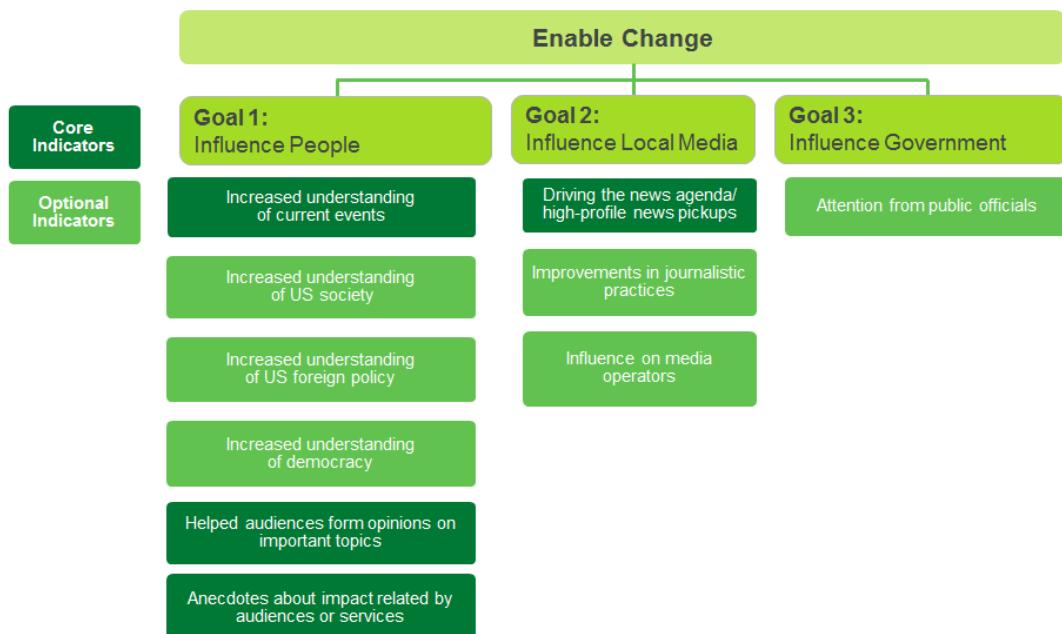
Thus, regularly evaluating audience attitudes about democracy, press freedom, internet freedom, corruption, governance, and self-efficacy on the BBG core country questionnaire, not just selectively or occasionally, would allow the BBG to access its impact on these fundamental drivers of democratization within and among countries over time. It would also allow the BBG to clearly align the attitudinal constructs that drive institutional change with contextual data from external sources such as: Freedom House indexes on democracy, press freedom, and Internet freedom; IREX Media Sustainability Index; World Governance Indicators; and the Failed States Index. Building credibility, trust, and engagement on these topics is the foundation upon which the BBG may expand its influence and impact on other topics, such as U.S. foreign policy, and thus comprehensively and consistently measuring BBG influence on these audience attitudes

Impact Framework



COMPONENT#1



COMPONENT #2**COMPONENT #3**

and perceptions is of great importance to the BBG Impact Framework in the long-term.

NOTES

1. Defining a BBG Impact Framework (2014).
2. Broadcasting Board of Governors 2013 Annual Report, p. 8.
3. BBG measures and reports unduplicated audience, meaning each individual, regardless of how many programs watched or listened to, is counted only once.
4. This design is akin to a 3-stage funnel: (1) start with the mouth of the funnel, in the form of large-N surveys to uncover interesting or important areas for further exploration; (2) then turn to the middle section of the funnel, in the form of medium-N studies, such as large scale sets of focus groups (akin to the Cuba study), controlled experiments, in-depth in-person survey interviews, etc., to begin focusing on these key areas; and (3) finally, for the narrowest part of the funnel, conduct focus group interviews based on the insights derived from the first two stages, with participants carefully recruited to reflect the key target demographics. In the reports, stage 2 is entirely missing from all but one study (Cuba), and in no cases were we persuaded—that is, Gallup made no effort to persuade the reader—that the small-N samples were apt (that is, representative of whatever would be the key target audiences).
5. For example, in Pakistan, in-depth interviews were evaluate the desires and expectations of radio programming and to examine of the role of mobile phones for news and information, among other things (*International Audience Research Project*. 2013. “Radio Mashaal In-Depth Interviews in Peshawar, Pakistan: 2012-2013 Qualitative Report”). Another study analyzed the media market in Vietnam, describing the state of access to information, and cable and satellite TV in particular (*International Audience Research Project*. 2013. “Television Platforms in Vietnam: 2013 Qualitative Research Study”).
6. For empirical evidence on this point, see: Goldman, Seth K., Diana C. Mutz, and Susanna Dilliplane (2014) All Virtue is Relative: A Response to Prior. *Political Communication* 30:4: 635-53; Prior, Markus (2014) The Challenge of Measuring Media Exposure: Reply to Dilliplane, Goldman, and Mutz. *Political Communication* 30:4: 620-634; Dilliplane, Susanna, Seth K. Goldman, and Diana C. Mutz (2013) Televised Exposure to Politics: New Measures for a Fragmented Media Environment. *American Journal of Political Science* 57:1: 236-248.
7. See for example Horne, Cale and Ryan Bakker. (2009) Public opinion in an autocratic regime: An analysis of Iranian public opinion data 2006-2008. Midwest Political Science Association 2009 Annual Convention, Chicago, Illinois, March 2009; Tsai, L., (2010). Quantitative research and issues of political sensitivity in rural China. In A. Carlson, M. Gallagher, K. Lieberthal, & M. Manion, (eds.) *Contemporary Chinese Politics New Sources, Methods, and Field Strategies*. New York: Cambridge University Press, pp. 246-265.
8. DeFleur, M. M. L., and DeFleur, M. H. (2003). *Learning to hate Americans: How US media shape negative attitudes among teenagers in twelve countries*. Spokane, WA: Marquette Books.
9. Eveland, W. P., Jr., & Cooper, K. E. (2013). An integrated model of communication influence on beliefs. *Proceedings of the National Academy of Sciences*. 110(3), 14088-14095 doi: 10.1073/pnas.1212742110.
10. To focus on one specific example, in the Iraq survey, Gallup weighted the 2012 data by gender, age, and education. They did not do this in the prior survey from 2011 (which was conducted by a different contractor). They then stipulate that this decision could account for differences between the surveys. We agree that it most likely contributes to such differences. We further understand from Gallup that they did not feel they had adequate information about the prior survey to prepare comparable weights. If true, this is unfortunate. Once again, this calls into at least some question the appropriateness of drawing such comparisons. All that said, if we took those trends at face value, then the question becomes, where did the audiences who left US broadcasters between 2011 and 2012 go? There's no clear way to tell from these data. We would think this would be the key follow-up question, again, assuming we believe the trend data.
11. This is similar to the Knowledge Networks Knowledge Panel (<http://www.knowledgenetworks.com/GANP/>), which provides the hardware in exchange for participants agreeing to use the technology and response to some number of snap polls).
12. Raudenbush, S.W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage; Snijders, T., & Bosker, R. J. (2003). *Multi-level analysis*. London: Sage Publications.; Nisbet E.C. (2006). The engagement model of opinion leadership: testing validity within a European Context. *International Journal of Public Opinion Research*, 18 (6), 3-30; Weimann, G. (1994). *The influentials*. Albany, NY: State University of New York Press. Weimann, G., Tustin, D. H., van Vuuren, D., & Joubert, J. P. R. (2007). Looking for opinionleaders: Traditional vs. modern measures in traditional societies. *International Journal of Public Opinion Research*, 19, 173-190.
13. Nisbet E.C. (2006). The engagement model of opinion leadership: testing validity within a European Context. *International Journal of Public Opinion Research*, 18 (6), 3-30.
14. See for example Mattes, R. & Bratton, M.(2007). Learning about democracy in Africa: Awareness, performance and experience. *American Journal of Political Science*, 51, 1, 192-217; Nisbet, E. C. (2008). Media use, democratic citizenship, and communication gaps in developing democracy. *International Journal of Public Opinion Research*, 20, 454-482; Nisbet, E. C., Stoycheff, E., & Pearce, K. E. (2012). Internet use and democratic demands: A multinational, multilevel model of Internet use and citizen attitudes about democracy. *Journal of Communication*, 62, 249–265; Welzel, C. (2007). Are levels of democracy affected by mass attitudes? Testing attainment and sustainment effects on democracy. *International Political Science Review*, 28, 397-424.
15. Nisbet, E.C. & Stoycheff, E. (2013). Let the people speak: a multi-level model of supply and demand for press freedom. *Communication Research*. 40(5), 720-741.

a. This review is based to our assessment of the 66 Gallup reports outlined in Appendix II, 4 documents related to the BBG's Impact Framework (including a sample survey instrument and its strategic outline), the description of BBG's FY2014 pilot panel research program, and interviews with researchers at BBG and Gallup. Additional research, including studies commissioned by Research Directors and the Office of Strategy and Development, were not included in this analysis.

File	Type	Fieldwork FY	Client(s)
BBG Impact Model Summary 02192014	N/A	N/A	N/A
Panels for Measurement of BBG's Impact	N/A	N/A	N/A
2014 Somalia Field (Impact Model)	Survey	2014	VOA
Operational Definitions (Impact Model)	N/A	N/A	N/A
2014 Nigerian Analytical Report.pdf	Survey	2014	VOA
2012 Afghanistan Focus Group Report - Karwan.pdf	FG	2013	VOA
2012 Afghanistan Focus Group Report - Radio Azadi.pdf	FG	2013	RFE-RL
2012 Azerbaijan Focus Group Report - Azadliq.pdf	FG	2013	RFE-RL
2012 Bangladesh Analytical Report.pdf	Survey	2012	VOA
2012 Burma Analytical Report.pdf	Survey	2012	RFA+VOA
2012 Burma In-Depth Interviews - RFA.pdf	IDIs	2012	RFA
2012 DRC Analytical Report.pdf	Survey	2013	VOA
2012 Egypt Analytical Report.pdf	Survey	2012	MBN
2012 Iran Analytical Report.pdf	Survey	2012	VOA
2012 Iran Focus Group Report.pdf	FG	2013	VOA
2013 BBG Iraq Analytical Report	Survey	2012	MBN
2012 Jordan Analytical Report.pdf	Survey	2013	MBN
2012 Kazakhstan Analytical Report.pdf	Survey	2013	RFE/RL
2012 Kazakhstan Focus Group Report.pdf	FG	2013	RFE/RL
2012 Kyrgyzstan Analytical Report.pdf	Survey	2013	RFE/RL
2012 Morocco Analytical Report.pdf	Survey	2012	MBN
2012 Moscow Focus Group Report.pdf	FG	2013	VOA
2012 Nigeria Analytical Report.pdf	Survey	2012	VOA
2012 Nigeria Mobile Phone Usage Focus Group Report.pdf	FG	2012	VOA
2012 North Korea Analytical Report - Traveler-Refugee-Defector.pdf	Survey	2013	VOA/RFA
2012 Russia Analytical Report.pdf	Survey	2012	VOA/RFE-RL
2012 Zimbabwe Analytical Report.pdf	Survey	2012	VOA
2012 Zimbabwe Focus Group Report - Satellite TV.pdf	FG	2013	VOA
2013 China In-Depth Interviews - Sensitive News (RFA).pdf	IDIs	2013	RFA
2013 Ethiopia Analytical Report.pdf	Survey	2013	VOA
2013 FATA Focus Group Report - Deewa.pdf	FG	2013	VOA
2013 Peshawar In-Depth Interviews - Mashaal.pdf	IDIs	2013	RFE-RL
2013 Russia Focus Group Report - Svoboda.pdf	FG	2013	RFE-RL
2013 Uyghur In-Depth Interviews.pdf	IDIs	2013	RFA
2013 Vietnam Analytical Report.pdf	Survey	2013	VOA/RFA
2013 Vietnam Qualitative Report - TV Platforms.pdf	IDIs	2013	VOA
Armenian FY2013 Armenia National Survey Sept 2012 Report-Final.pdf	Survey	2013	VOA/RFE-RL
Azerbaijan National Survey Report Sept 2012 - FINAL FY2013 Azerbaijani.pdf	Survey	2013	VOA/RFE-RL
Bangui 2012 survey report FINAL.pdf	Survey	2013	VOA
BBG 2012 Ghana FINAL FG Report.pdf	FG	2012	VOA

BBG Rwanda (Central Africa) FINAL Focus Group Report.pdf	FG	2013	VOA
BBG_2012_Survey_Rwanda_Final_Analytic_Report.pdf	Survey	2012	VOA
BBG_2012_Cambodia_Analytic_Report_Final.pdf	Survey	2013	VOA/RFA
BBG_2012_DRC_Analytic_Report_Final.docx	Survey	2013	VOA
BBG_2012_ETHIOPIA_Analytical_Report_Final_103013.pdf	Survey	2013	VOA
BBG_2012_Indonesia_Final_Analytical_Report.pdf	Survey	2012	VOA
BBG_2012_Laos_Analytical_Report_FINAL.pdf	Survey	2013	VOA/RFA
BBG_2012_Somalia_Analytic_Report_Final_Survey.pdf	Survey	2013	VOA
BBG_2012_Uganda_Analytic_Survey_Report_Final.pdf	Survey	2012	VOA
BBG_2012_Uzbekistan_Analytic_Report_Final.pdf	Survey	2012	VOA/RFE-RL
BBG_Bamako_2012_Analytic_Report_Final.pdf	Survey	2013	VOA
BBG_Liberia_2012_Analytic_Report_Final.pdf	Survey	2012	VOA
China Satellite FINAL Report 03 26 2013.pdf	IDIs	2013	VOA
Haiti National Survey July 2012 - Analytical Report October 2012-FINAL.pdf	Survey	2012	VOA
Lavvadamerik VOA Creole Focus Groups Report Sept 2012-FINAL.pdf	FG	2012	VOA
Georgian FY2013 Georgia National Survey Sep-Oct 2012 Report FINAL.pdf	Survey	2013	VOA/RFE-RL
International TV News Providers- Focus Groups With Heavy News Viewers in Quito and Lima Dec 2012 Report FY2013 FINAL.pdf	FG	2013	VOA
Analytical Summary of Cuban Immigrant Focus Group Research Jun 2012-Jan 2013 - FINAL.pdf	FG	2013	OCB
Analytical Summary of Media Use Questionnaire Among Cuban Immigrants Jun 2012-Jan 2013 - FINAL.pdf	FG	2013	OCB
Radio Marti Listener Panel - June-July 2012 - Main Report FINAL.pdf	Panel	2012	OCB
TV Marti Viewer Panel - December 2012 - Main Report FINAL.pdf	Panel	2013	OCB
Pakistan NS 2012 Final.pdf	Survey	2013	VOA
Pakistan Sana FG Report Final.pdf	FG	2013	VOA
Pakistan Zindeggi 360 FINAL FG Report.pdf	FG	2013	VOA
Portuguese to Africa FY2013 VOA Portuguese Program Vida sem medo - Mozambique Qualitative - March 2013 - FINAL Report.pdf	FG	2013	VOA
Swahili BBG2012 Tanzania Survey Analytical Report.pdf	Survey	2012	SOV
Ukrainian FY2013 Holos Ameryky TV and New Media Focus Groups Kyiv-Kharkiv Nov 2012 Report FINAL.pdf	FG	2013	VOA
Ukrainian FY2013 Ukraine National Survey Oct 2012 Report-FINAL.pdf	Survey	2013	VOA/RFE-RL
Venezuela Urban Survey Dec 2012 Analytical_Report_FINAL.pdf	Survey	2013	VOA
VOA in the Sahel_WP data_final.pdf	Survey	2013	VOA

b. For example, see *International Audience Research Project* (2013). "Radio Free Asia In-depth Interviews With Uyghurs in Istanbul, Turkey: 2013 Qualitative Report."

c. For examples of using longitudinal, cross-country data collected consistently over several years to evaluate causality, see: Welzel, C. (2007). Are levels of democracy affected by mass attitudes? Testing attainment and sustainment effects on democracy. *International Political Science Review*. 28, 397-424. doi:10.1177/0192512107079640; Mattes, R., & Bratton, M. (2007). Learning about democracy in Africa: Awareness, performance and experience. *American Journal*

d. There were a few examples where the reports central findings seemed to contradict data that was later described in detail. For example, a report examining the significance of RFA's Uyghur programming claimed, "RFA remains a go-to source for news and information of the most sensitive kind." But, in the section of the report detailing perceptions of RFA, it was clear that a large number of the interviewees (who were chosen because they expressed interest in accessing sensitive news) were unfamiliar with RFA: "...only a few participants say they have accessed the RFA website prior to participating in this study" (see *International Audience Research Project* (2013). "China Sensitive News In-Depth Interviews," p. 4, 27). While this discrepancy is hard to explain, it raises another fundamental question: why didn't the report specify how many of the 24 research subjects knew about RFA prior to the interviews? In addition, in its report on TV offerings and consumption in Vietnam, Gallup found that "Users rarely watch news programming aside from VTV news. Vietnamese seem to be highly satisfied with the news programming provided by state-operated VTV" but still went on to suggest that "VOA should create its own 24/7 channel" (*International Audience Research Project* (2013). "Television Platforms in Vietnam: 2013 Qualitative Research Study," p. 28; 33.) In a similar report on satellite TV use in China, Gallup found that VOA Mandarin and RFA were already accessible in some households, but that none of the research participants expressed any interest in accessing either channel (or any foreign news content) (From: *International Audience*

Research Project (2013). "Voice of America and the Satellite Television Landscape in China: 2012 Qualitative Research Study," p. 18). Despite this, Gallup recommended "VOA conduct further quantitative research into satellite TV and its alternatives to identify more appropriate media for broadcasting in China," adding, "Additionally, VOA may want to consider using alternative media types to broadcast its programming such as the Internet TV boxes and smartphone technology" (*International Audience Research Project* (2013). "Voice of America and the Satellite Television Landscape in China: 2012 Qualitative Research Study," p. 7). Such recommendations—calling for research into additional ways of distributing content that there is little demand for—seem to lack a basic understanding of the role international broadcasting, or under what circumstances international broadcasters can be effective. A third example, again from the 2013 report on Uyghurs perceptions of RFA, raises some questions regarding how data are interpreted and findings determined. Based on 12 IDIs with Uyghurs who had traveled to Istanbul, Turkey, the report argued that RFA was "the main, if not sole, provider of comprehensive Uyghur-related news and information, particularly via its website." But the report also states that "Participants note that receiving RFA Uyghur content is challenging because its website is blocked and radio signals are jammed. The same report goes on to suggest that "Facebook is an important tool for accessing and sharing information and resources among the Uyghur community. RFA should take advantage of the opportunity to engage this online community of Uyghurs by building a strong presence on its network" despite acknowledging that "Facebook is blocked in XUAR and its mobile app cannot be downloaded," as well as "News and information sharing beyond every day conversational topics is limited, if not non-existent between Uyghurs abroad and Uyghurs in the XUAR since the July 2009 unrest." If Facebook isn't accessible in XUAR, and it is Uyghurs outside of China are not sharing sensitive information with friends and family in the country, then prioritizing RFA's Facebook presence would seem to do little to achieve the stated goal of providing "accurate news and information to Uyghurs residing in the Uyghur region" (*International Audience Research Project* (2013). "Radio Free Asia In-depth Interviews With Uyghurs in Istanbul, Turkey: 2013 Qualitative Report," p. 5, 6, 8, 11 and 17).

CONCLUSION

Public diplomacy officers should be proud of their lead within the State Department and embrace the opportunity further to improve their approach to research and evaluation. They also have an opportunity to be an interagency leader on research and evaluation when it comes to foreign public engagement. Likewise, the Broadcasting Board of Governors has made profound strides in improving the strategic focus and efficiency of its worldwide activities through its pending reorganization and its creation of an Impact Framework to understand how its global programming advances the U.S. foreign policy objectives of supporting freedom and democracy. We look forward to seeing how this progresses.

While completing this report, many indicated a seeming lack of interest among State Department and BBG officials in research for planning and program design. Even when data is produced, it is not actionable and incorporated into the workflow. A “culture of research” in U.S.

public diplomacy and international broadcasting would recognize the value of research-based strategic planning, provide benchmarks for measurement, and move toward more systematically understanding the impact of programs to inform and engage foreign audiences.

Overall, we found meaningful movement toward attaching research and evaluation of public diplomacy and international broadcasting activities to U.S. foreign policy goals. Much of the success of research and impact evaluation depends on leadership that consistently signals that data-driven strategies and tactics are important. State Department and BBG leadership should continue to work to institutionalize processes that integrate research at the outset of planning campaigns and programs. Yet this also depends on significant structural and organizational change, not the least of which is increased funding closer to the 5 percent industry standard and the employment of more specialists.

SUMMARY OF RECOMMENDATIONS

STRUCTURAL & ORGANIZATIONAL CHANGE

1. Create a Director of Research and Evaluation Measurement Position & Expand Evaluation and Measurement Unit in Policy, Planning and Resources (R/PPR/EMU)
2. Support Evaluation Staff with More Expertise
3. Increase Funding for Research and Evaluation
4. Review Further The Privacy Act of 1974
5. Provide State Department Research Waiver for the Paperwork Reduction Act of 1980
6. Improve Inter- and Intra-Agency Cooperation and Data Sharing
7. Support a Risk-Taking Culture That Allows for PD and Broadcasting Setbacks
8. Establish Guidance and Training on Research and Evaluation
9. Create ACPD Subcommittee on Research and Evaluation

METHODS AND RESEARCH DESIGN

1. Acknowledge Research Limitations in Products
2. Increase Integration of Data into Strategy and Program Development
3. Create More Disaggregated Data for Better Feedback Loops
4. Use More Comparative Data and/or Analysis to Determine Impact
5. Provide More Contextual Data to Determine Impact
6. Highlight Negative Findings for Couse Correction

For an electronic version of the report,
please visit <http://www.state.gov/pdcommission/>
or scan the QR Code below.



September 16, 2014